МЕДИЦИНСКИ УНИВЕРСИТЕТ – СОФИЯ
МЕДИЦИНСКИ ФАКУЛТЕТ

КАТЕДРА ПО МЕДИЦИНСКА ГЕНЕТИКА
РЪКОВОДИТЕЛ: ПРОФ. Д-Р ДРАГА ТОНЧЕВА, ДБН

СЕНА КАРАЧАНАК-ЯНКОВА

ГЕНЕТИЧНА СТРУКТУРА НА БЪЛГАРСКАТА ПОПУЛАЦИЯ ПО Y-ХРОМОЗОМАТА И МИТОХОНДРИАЛНАТА ДНК

А В Т О Р Е Ф Е Р А Т

на дисертационен труд за присъждане на
образователна и научна степен "ДОКТОР"

Област на висше образование: 4. „Природни науки, математика и информатика"
Професионално направление: 4.3. „Биологически науки"
Докторска програма: „Генетика"

НАУЧЕН РЪКОВОДИТЕЛ: ПРОФ. Д-Р ДРАГА ТОНЧЕВА, ДБН

София, 2013

Изследванията по дисертационния труд са финансирани частично от Фонд „Научни изследвания” към МОМН по научен проект „Характеризиране на антропо-генетичната идентичност на българския народ”, договор № ДОО 2-110/22 май 2009 г.

Изследванията са извършени в: Катедрата по Медицинска генетика, Медицински факултет, Медицински университет-София; Катедрата по биология и биотехнологии "L. Spallanzani" към Университета в Павия, Италия и лабораторията на Проген ООД, София.

Докторантът е зачислен на самостоятелна подготовка към Катедрата по Медицинска генетика, Медицински факултет, Медицински университет – София.

Дисертационният труд съдържа 190 страници и е онагледен с 25 фигури, 21 таблици и 2 приложения. Библиографският списък обхваща 280 литературни източника, от които 220 са от последните 10 години.

Дисертационният труд е обсъден, одобрен и насочен за защита на заседание на разширен катедрен съвет на Катедрата по Медицинска генетика, Медицински факултет, Медицински университет – София на 03 юни 2013 г.

Публичната защита ще се състои на 10 септември 2013 г. от 15.00 часа в аудиторията на СБАЛАГ „Майчин дом” на основание Заповед № РК36-1290/ 24.06.2013 г. на Ректора на МУ – София, пред научно жури в състав:

Председател:

Проф. д-р Драга Иванова Тончева, дбн - становище

Членове:

Чл.-кор. проф. Иван Георгиев Иванов, дбн – рецензент

Проф. Спаска Димитрова Петкова, дбн – рецензент

Проф. Севдалин Георгиев Ангелов, дбн – становище

Доц. Елена Георгиева Тодоровска, дб - становище

Материалите по защитата са на разположение в Катедрата по Медицинска генетика, Медицински университет – София, 6^{ти} етаж, СБАЛАГ „Майчин дом”, ул. Здраве № 2, 1431 София.

СЪДЪРЖАНИЕ

Списък на по –често използвани съкращения 4
ВЪВЕДЕНИЕ 5
ЦЕЛ И ЗАДАЧИ 6
МАТЕРИАЛИ И МЕТОДИ 7
1. Участници в проучването 7
2. Биологичен материал 7
3. Генотипиране на мтДНК 7
3.1. Секвениране на контролния регион на мтДНК 8
3.2. PCR-RFLP анализ на маркери в кодиращия регион на мтДНК 8
4. Генотипиране на Y-хромозомата 10
4.1. Анализ на биалелни маркери 10
4.2. Анализ на микросателити 11
5. Статистически анализи 11
5.1. Вътрепопулационно сравнение 11
5.2. Междупопулационно сравнение 11
5.3. Анализ на микросателитната изменчивост в Y-хромозомни хаплогрупи 12
РЕЗУЛТАТИ 13
1. Определяне на мтДНК хаплотипове и хаплогрупи при българите 13
1.1.Честота на митохондриалните хаплогрупи при българите 14
1.3. Междупопулационно сравнение на честотата на мтДНК хаплогрупи 17
2. Определяне на Y-хромозомни хаплогрупи и хаплотипове при българите 20
2.1. Честота на Y-хромозомните хаплогрупи при българите 20
2.2. Структура на българския Y-хромозомен генофонд в страната 24
2.3. Междупопулационно сравнение на честотата на Y-хромозомни хаплогрупи 25
2.4. Y-STR хаплотипове на чести Y-хромозомни хаплогрупи 30
3. Генетична структура на европейски популации въз основа на SNPs в целия геном 34
ОБСЪЖДАНЕ 34
Обсъждане на генетичната изменчивост на българите от гледна точка на мтДНК 34
Обсъждане на генетичната изменчивост на българите от гледна точка на Y-хромозомата 39
Обсъждане на резултатите от геномния анализ на SNPs 46
ЗАКЛЮЧЕНИЕ 48
ИЗВОДИ 50
НАУЧНИ ПРИНОСИ 51
ЛИТЕРАТУРНИ ИЗТОЧНИЦИ 52
НАУЧНИ ПУБЛИКАЦИИ И СЪОБЩЕНИЯ ПО ТЕМАТА 54
БЛАГОДАРНОСТИ 56
SUMMARY 57

Забележка: Номерата на фигурите и таблиците в автореферата не съотвестват на номерата в дисертационния труд.

Списък на по –често използвани съкращения

Съкращения на кирилица:

ДНК - дезокси рибонуклеинова киселина
мтДНК - митохондриална ДНК
Хг – хаплогрупа

Съкращения на латиница:

AMOVA-Analysis of Molecular Variance (анализ на молекулната дисперсия)
bp - base pair (базова двойка)
CRS - Cambridge Reference Sequence (кеймбриджка референтна последователност)
DHPLC - Denaturing High Performance Liquid Chromatography (денатурираща високо изпълнителна течна хроматография)
EDTA - Ethylene Diamine Tetraacetic Acid (етилендиаминтетраоцетна киселина)
HVS-I– Hypervariable Segment I (първ хипервариабилен сегмент)
HVS II- Hypervariable Segment II (втор хипервариабилен сегмент)
kya – kilo years ago (хиляди години преди)
np – nucleotide position (нуклеотидна позиция)
PC – Principal Component (главен компонент)
PCA- Principal Component Analysis (анализ на главните компоненти)
PCR - Polymerase Chain Reaction (полимеразна верижна реакция)
RFLP – Restriction Fragment Length Polymorphism (полиморфизъм по дължината на рестрикционните фрагменти)
SNP – Single Nucleotide Polymorphism (единичен нуклеотиден полиморфизъм)
STR –Short Tandem Repeat (кратък тандемен повтор)
YAP -Y Alu polymorphism (Y-хромозомен Alu полиморфизъм)
Y-STR – Y-chromosome Short Tandem Repeat (кратък тандемен повтор на Y-хромозомата)

ВЪВЕДЕНИЕ

Генетичната структура на популациите и процесите, които я поддържат и променят са предмет на изследване на популационната генетика. Човешката популационна генетика намира широко приложение в различни сфери на науката, включително и в медицинската генетика, където подходите й се използват за разкриване на генетичните причини, водещи или предразполагащи към наследствени болести.

Популационно-генетичните изследвания имат дългогодишна традиция, започваща с анализ на фенотипни характеристики с менделов тип на унаследяване (кръвни групи, алозими и пр.). С усъвършенстването на молекулните техники, популационните изследвания се насочват към анализ на генетични полиморфизми.

Обектът на популационно-генетичните изследвания – генетичната изменчивост на съвременните популации всъщност представлява съкровищница на събитията от тяхното близко и далечно генетично минало. Един от най-ефективните подходи за вникване в това минало е изследването на определени варианти в митохондриалната ДНК (мтДНК) и Y-хромозомата. Това е възможно благодарение на техните отличителни характеристики, като унаследяването на мтДНК само от майката и предаването на Y-хромозомата в потомството само от баща на син. Еднородителското унаследяване и липсата на рекомбинация в мтДНК и 95% от Y-хромозомата показват, че разликите в техните последователности се дължат единствено на последователното натрупване на промени (мутации) през миналото. Изследването на натрупаната по този начин изменчивост на мтДНК и Y-хромозомата спомага за проследяване на произхода по майчина и бащина линия, съответно.

За разлика от останалите страни в Европа, България изостава в изследванията на изменчивостта на еднородителските системи. До настоящото проучване, в научната литература данните за мтДНК и Y-хромозомната изменчивост на българите са само от стари и недостатъчно задълбочени изследвания. Освен това броят на изследваните българи е малък и получените резултати не са обсъждани от историческа гледна точка.

За да се запълни този сериозен пропуск за страната и за Европа като цяло, настоящото проучване цели да характеризира генетичната структура на българската популация по мтДНК и Y-хромозомата и да включи българите в изследванията на цялостната геномна структура в европейски контекст. Изследването на мтДНК и Y-хромозомната изменчивост на българската популация е проведено върху голяма и представителна за страната извадка, като е приложен подход, с който се достига високо ниво на филогенетична резолюция. Генетичните резултати са анализирани на микро- и макрогеографско ниво и установените профили са обсъдени в

исторически контекст.

Настоящото проучване е първото по рода всеобхватно съвременно изследване на генетичната структура на българската популация. То предоставя възможност за определяне на генетичното родство и различия на българите с други европейски и неевропейски популации.

ЦЕЛ И ЗАДАЧИ

Цел

Да се определи хаплогрупният състав на съвременните българи по мтДНК и Y-хромозомата на макро- и микро-географско ниво.

Задачи

1. Създаване на представителна извадка от ДНК проби за цялата страна;
2. Установяване на честотата на мтДНК хаплогрупите (Хг) в българската популация;
3. Определяне на разпределението на мтДНК хаплогрупите в страната;
4. Анализиране на родствените връзки по майчина линия на българите с други популации;
5. Установяване на честотата на Y-хромозомните хаплогрупи в българската популация;
6. Определяне на разпределението на Y-хромозомните хаплогрупи в страната;
7. Анализ на родствените връзки по бащина линия на българите с други популации;
8. Проследяване на времето и мястото на възникване на микросателитната изменчивост на чести Y-хромозомни хаплогрупи;
9. Сравнение на българите с други европейски популации на цялостно геномно ниво въз основа на данни за единични нуклеотидни полиморфизми (SNPs);
10. Интерпретация на генетичната структура на съвременните българи от историческа гледна точка.

МАТЕРИАЛИ И МЕТОДИ

1. Участници в проучването

Благодарение на съдействието на началника на Военномедицинска академия (ВМА) – проф. Стоян Тонев, събирането на участници за проучването е проведено в Центъра по трансфузионна хематология към ВМА, ръководен от д-р Румен Попов. Сред кърводарителите са подбрани неродствени лица от български произход. След запознаване с изследването и кратко интервю от участниците е взето информирано съгласие.

Изследването на мтДНК генофонда на съвременните българи е проведено върху извадка от 855 мъже. Анализът на Y-хромозомното разнообразие на българите е извършен върху 808 лица, които са част от участниците в мтДНК изследването. Въз основа на предоставената информация, участниците в проучването са класифицирани към една от областите на страната. При мтДНК анализа класифицирането е въз основа на месторождението на майката, което беше прецизно съобщено от 814 от участниците; а при Y-хромозомата - според месторождението на бащата, уточнено за 739 от участниците.

Освен за мтДНК и Y-хромозомния анализ са подбрани и 47 неродствени лица от двата пола, които са включени в европейско популационно-генетично проучване с цялостен геномен анализ на над 270 000 единични нуклеотидни полиморфизми (SNPs).

2. Биологичен материал

От всеки участник са взети 10 ml кръв във вакутейнери с антикоагулант K_3EDTA (калиева сол на EDTA) при спазване на стандартните процедури за стерилност. Кръвта е съхранявана на +4ºC до транспортирането в молекулярно-генетичната лаборатория в Катедрата по Медицинска генетика, Медицински университет-София, където изолирането на високомолекулна ДНК е извършвано не по-късно от 24 часа след вземането на кръвта.

Изолирането на ДНК от кръвните проби е извършено чрез фенол-хлороформна екстракция, последвана от етанолна преципитация.

3. Генотипиране на мтДНК

Генотипирането на мтДНК проведохме в лабораторията на проф. Антонио Торони към Катедрата по биология и биотехнологии "L. Spallanzani" в Университета в Павия, Италия. В настоящия труд при класифицирането към мтДНК (под-) хаплогрупи използвахме два метода за генотипиране: ДНК секвениране и анализ на полиморфизми по дължината на рестрикционните фрагменти (RFLPs). На секвениране е подложена голяма част от контролния регион, която обхваща целия първ (HVS-I) и част от втория хипервариабилен

сегмент (HVS-II). Хаплотиповете на контролния регион, установени от секвенирането, са групирани в определени мтДНК (под-) хаплогрупи въз основа на наличието на маркери в HVS-I и -II. Предполагаемата хаплогрупна принадлежност е потвърдена чрез филогенетично йерархичен RFLP анализ на маркери в кодиращия регион на мтДНК. Класифицирането към мтДНК (под-) хаплогрупи следва филогенетичните взаимоотношения и номенклатурата, описани от van Oven & Kayser, 2009.

3.1. Секвениране на контролния регион на мтДНК

На автоматично секвениране по Sanger са подложени успешно амплифицирани продукти на контролния регион на мтДНК.

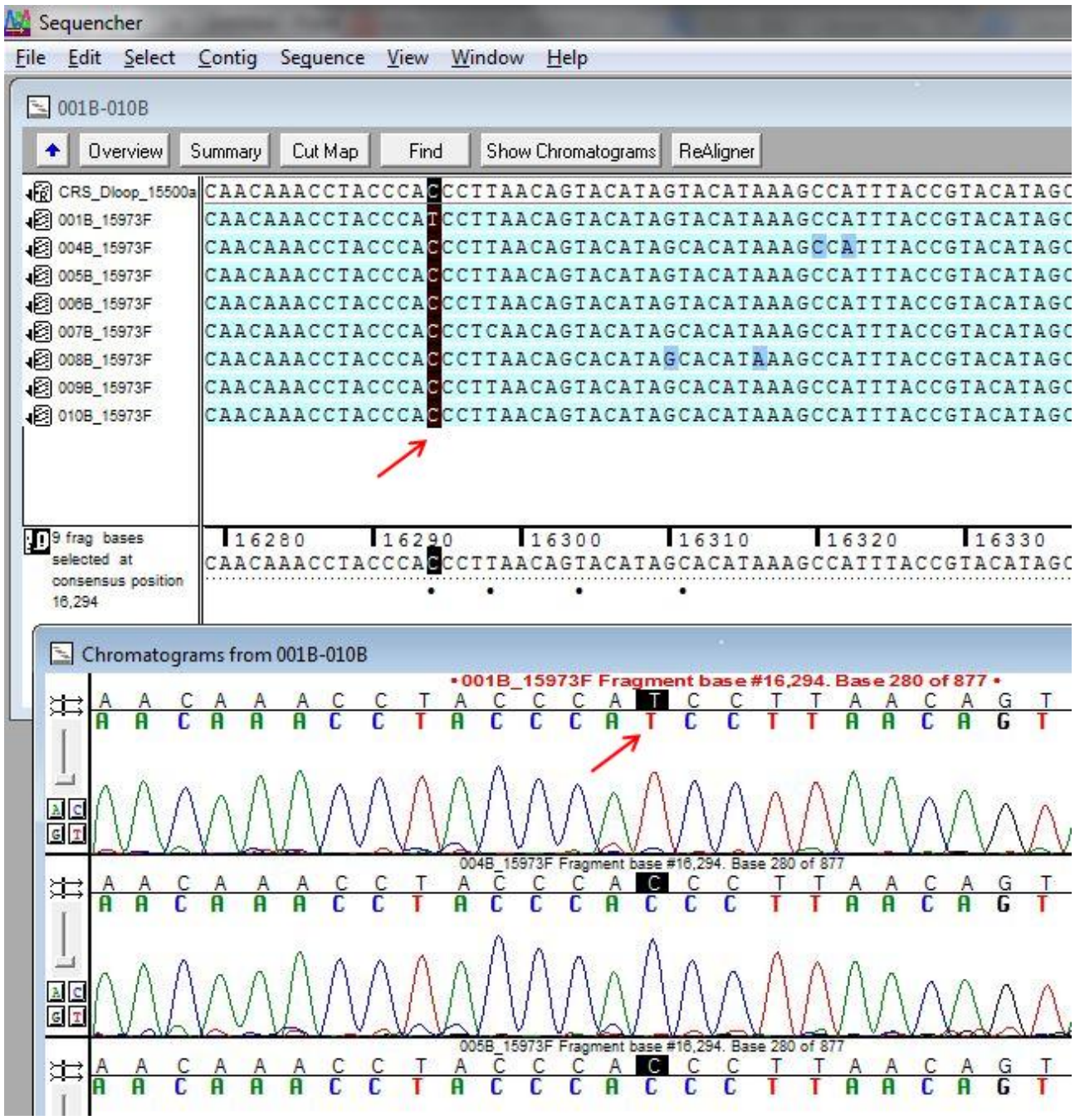
Допълнителни технически детайли са достъпни на сайта на Центъра по интердисциплинарно проучване чрез иновативни биотехнологии CRIBI (Centro Ricerche Interdipartimentale Biotecnologie Innovative) към Университета в Падуа: http://www.bmr-genomics.it/seq_i-index.html.

Последователностите на контролния регион на мтДНК са сравнени с тези на кеймбриджката референтна последователност (CRS) с помощта на софтуер Sequencher v4.8 (Gene Codes Corp.) (Фигура 1).

Всички последователности са анализирани от две независими лица и позициите на установените разлики със CRS (полиморфизмите) са нанесени в таблица. След това, наблюдаваните хаплотипове на контролния регион са причислени към мтДНК филогенетични линии въз основа на маркери в HVS-I и -II, които са отличителни и/или свързани с определени мтДНК (под-) хаплогрупи.

3.2. PCR-RFLP анализ на маркери в кодиращия регион на мтДНК

Хаплогрупната принадлежност на хаплотиповете е потвърдена чрез йерархичен анализ на RFLP маркери в кодиращия регион. За да се достигне висока филогенетична разделителна способност, в настоящото проучване при PCR-RFLP анализа са използвани 22 рестрикционни ензима.



Фигура 1. Секвенционни хроматограми получени чрез програмата Sequencer v4.8. Стрелките сочат транзицията (C -> T) в позиция 16 294 при проба 001B.

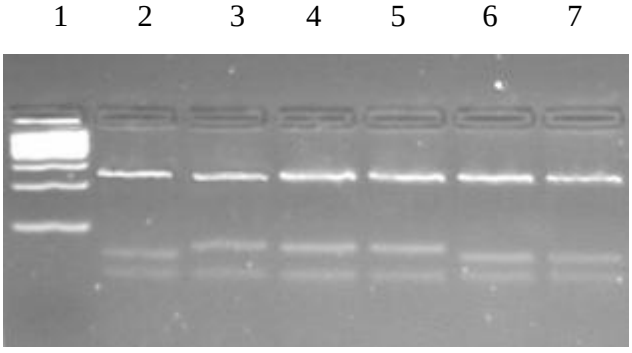
4. Генотипиране на Y-хромозомата

4.1. Анализ на биалелни маркери

Генотипирането на Y-хромозомата по отношение на биалелни маркери проведохме в лабораторията на проф. Орнелла Семино към Катедрата по биология и биотехнологии "L. Spallanzani" в Университета в Павия, Италия. За установяване на Y-хромозомните хаплогрупи, които се срещат при българите, са генотипирани 75 биалелни маркера. Първите анализирани маркери, които са изследвани във всички проби са M89 и M9, а останалите са изследвани в йерархична последователност, според филогенитичното дърво на Y-хромозомата. Хаплогрупите са означени в съответствие с препоръките на Y-хромозомния консорциум и според номенклатурата, основаваща се на мутациите.

В зависимост от изследвания маркер, при генотипирането на Y-хромозомата са използвани един от следните методи:

- определяне дължината на PCR продукта (за изследване на маркер YAP)
- секвениране (за маркери SRY4064/M96, P16, M304 и M73)
- PCR-RFLP анализ (за маркери M94, M2, M78, V13, M81, M123, M34, P15, L91, M547, P303, U1, M527, M406, Page19, M253, P37, M423, Page08, M172, Page55, M67, M12, M9, M61, M74, M242, M207, M198, M458, M269, L23, M412, U106, S116, U152, M529 и M70) (Фигура 2)
- денатурираща високо изпълнителна течна хроматография - DHPLC (Denaturing High Performance Liquid Chromatography) (за маркери M216, M217, M35, M89, M201, L497, M69, M82, M429, M170, M223, M267, M365, M390, M367, M368, M410, M530, M47, M92, M241, M214, M231, M120, M25, M346, M323, M378, M173, M343, L11 и M124).



Фигура 2. Електрофоретичен профил на рестрикционните фрагменти получени при RFLP анализ на маркер M269. При наличие на маркера продуктите от смилането с ензима *Tsp509*I са с размери 251, 42, 61, 10 и 5 bp, а при липсата на мутацията съответните фрагменти са с дължина 251, 42, 71, 10 и 5 bp. Старт 1: молекулен маркер за определяне концентрацията и размера на PCR фрагменти от 100 – 1 000 bp (100 bp DNA Ladder); стартове 2, 6 и 7: проби носещи маркера; стартове 3, 4 и 5: проби отрицателни по маркер M269.

4.2. Анализ на микросателити

Микросателитите на Y-хромозомата изследвахме в лабораторията на Проген ООД, София. Анализите са проведени върху 247 от изследваните проби, избрани според принадлежността към една от хаплогрупите E-V13, I-M423, J-M241, R-M458, R-L23*, R-U106 или R-U152. При Y-хромозомния микросателитен анализ е използван кит AmpFlSTR Yfiler PCR Amplification Kit (Applied Byosystems). Получените PCR продукти са изследвани на автоматичен секвенатор модел ABI Prism 310 Sequence Genetic Analyzer чрез фрагментен анализ. Резултатите са анализирани с помощта на софтуер GeneMapperID v3.2.1.

5. Статистически анализи

Честотите на наблюдаваните мтДНК и Y-хромозомни хаплогрупи са използвани за изследване структурата на генетичната изменчивост в страната (вътрепопулационно сравнение), както и за определяне на генетичното родство на българите с предишно изследвани популации (междупопулационни сравнения). Получените Y-STR резултати, заедно със съответни данни за други народи/райони, са използвани за датиране и проследяване на разпространението на определени Y-хромозомни хаплогрупи.

5.1. Вътрепопулационно сравнение

5.1.1. χ^2 (хи-квадрат) тест

В настоящото изследване чрез χ^2 тест е изследвано митохондриалното разнообразие вътре в България (т.е. разпространението на мтДНК хаплогрупите по районите на страната). Тестът е извършен с помощта на софтуер Excel, допълнен с XLSTAT.

5.1.2. Анализ на молекулната дисперсия - AMOVA (Analysis of Molecular Variance)

Рапределението на Y-хромозомната изменчивост в страната е оценена чрез AMOVA анализ при използване на софтуер Arlequin 3.5.

5.2. Междупопулационно сравнение

5.2.1. Анализ на главните компоненти – PCA (Principal Component Analysis)

За да се илюстрират родствените връзки на българите с други популации по майчина и по бащина линия е проведен анализ по метода на главните компоненти въз основа на честотата на мтДНК и Y-хромозомните хаплогрупи. Анализът на главните компоненти е извършен със софтуер Excel, допълнен с XLSTAT.

5.3. Анализ на микросателитната изменчивост в Y-хромозомни хаплогрупи

5.3.1. Датиране на Y-хромозомни хаплогрупи

Възрастта на Y-хромозомните хаплогрупи може да се определи чрез тяхната микросателитна изменчивост и чрез използване на определена мутационна скорост. Широко прието е датирането на Y-хромозомните хаплогрупи да се извършва в тези народи/райони, в които съответната хаплогрупа е представена с повече от пет проби. При оценката на възрастта на хаплогрупите в България е използвано географското разделяне на страната, за което AMOVA анализът показва наличието на значима междугрупова дисперсия.

Горната граница на времето на експанзия на хаплогрупите се определя според метода на Zhivotovsky et al., 2004 с промените от Sengupta et al., 2006. Недостатъчното познаване на някои събития от миналото (например: флуктуации в размера на популациите, популационни стеснения и т.н.) поражда определени отклонения в настройката на Y-STR молекулния часовник. Поради това изчислените датировки на микросателитната изменчивост трябва да се обмислят внимателно.

5.3.2. Мрежи на Y-хромозомните хаплогрупи

Филогенетичните връзки между Y-STR хаплотиповете в хаплогрупите I-M423, E-V13, R-M458, R-L23* и J-M241 се изобразяват чрез медиана-свързващи - MJ (Median-Joining) мрежи [Bandelt et al., 1999]. Те са конструирани чрез изполване на програмата Network 4.5.1.6 (www.fluxus-engineering.com). MJ мрежите се изчисляват чрез обработка на данните за броя на повторите с помощта на редуцираният медианен –RM (Reduced Median) метод [Bandelt et al., 1995], като на всеки STR локус се приписва тежест според обратно пропорционалната стойност на дисперсията на броя на повторите.

РЕЗУЛТАТИ

1. Определяне на мтДНК хаплотипове и хаплогрупи при българите

Секвениран е митохондриалния контролен регион на 855-те ДНК проби с дължина от около 750-800 bp: започващ от нуклеотидна позиция 16020 и завършващ до нуклеотидна позиция 200-300.

Изследваните 855 последователности на мтДНК контролния регион са групирани в 557 хаплотипа. Въз основа на разликите със CRS и на резултатите от PCR/RFLP анализа, хаплотиповете са класифицирани в 45 мтДНК филогенетично крайни хаплогрупи и подхаплогрупи.

Хаплотиповете, които се срещат с честота над 0.5% в цялата извадка са дадени в Таблица 1

Таблица 1. Най-чести хаплотипове на мтДНК контролния регион при българите, определени за последователността от нуклеотидна позиция (np) 16020 до нуклеотидна позиция 200 (прочетена за 850 ДНК проби от извадката).

Брой лица (%)	**Мутационен мотив (от np 16020 до np 200)**	**Хаплогрупа (%)**
46 (5.4)	16519	H (3.3), H1 (1.8), H3 (0.3)
12 (1.4)	16304	H5 (1.4)
11 (1.3)	16311	HV (0.7), H (0.6)
10 (1.2)	16069 16126 16366 16519 73 185 188	J1c (1.2)
9 (1.1)	16311 16519	H (0.5), HV (0.4), H1 (0.1), H3 (0.1)
9 (1.1)	16126 16163 16186 16189 16294 16519 73	T1 (1.1)
9 (1.1)	CRS	H (1.1)
7 (0.8)	16224 16311 16519 73 146 152	K1 (0.5), K2 (0.3)
7 (0.8)	16356 16519 73 195	U4 (0.8)
7 (0.8)	16287 16519 195	H (0.8)
6 (0.7)	16298 16311 72	HV0 (0.7)
6 (0.7)	16126 16287 16294 16519 73	T2 (0.7)
6 (0.7)	16362 16482	H6 (0.7)
6 (0.7)	16519 152	H (0.2), H1 (0.2), H3 (0.2)
6 (0.7)	16298 72	V (0.6), HV0a (0.1)
5 (0.6)	16343 73 150	U3 (0.6)
5 (0.6)	16172 16223 16231 16292 16519 73 189 194	W (0.6)
5 (0.6)	16162 16519 73	H (0.6)
5 (0.6)	16256 16270 16399 73	U5a (0.6)
5 (0.6)	16069 16126 185	J1c (0.6)
5 (0.6)	16069 16126 16261 73	J1c (0.6)

1.1.Честота на митохондриалните хаплогрупи при българите

Честотата на 45-те филогенетично крайни мтДНК хаплогрупи при българите и честотата на основните хаплогрупи или макрохаплогрупи, към които те принадлежат са сумирани в Таблица 2.

Таблица 2. Честота на основните мтДНК хаплогрупи/макрохаплогрупи и техните подхаплогрупи при българите.

Хаплогрупа	**Подхаплогрупа**	**Брой (%)**
H:		359 (42.0)
	H*	222 (26.0)
	H1	73 (8.5)
	H3	8 (0.9)
	H5	29 (3.4)
	H6	14 (1.6)
	H8	8 (0.9)
	H11	5 (0.6)
HV0:		32 (3.7)
	HV0*	13 (1.5)
	HV0a	6 (0.7)
	V	13 (1.5)
HV(xH,HV0):		33 (3.9)
	HV*	30 (3.5)
	HV1	2 (0.2)
	HV2	1 (0.1)
R:		7 (0.8)
	R*	3 (0.4)
	R0a	4 (0.5)
U:		188 (22)
	U1	12 (1.4)
	U2	12 (1.4)
	U3	16 (1.9)
	U4	33 (3.9)
	U5	57 (6.7)
	>U5a	37 (4.3)
	>U5b	20 (2.3)
	U7	4 (0.5)
	U8	4 (0.4)
	>U8a	2 (0.2)
	>U8b	2 (0.2)
	K	50 (5.9)
	>K1	45 (5.3)
	>K2	5 (0.6)

продължена

Хаплогрупа	Подхаплогрупа	Брой (%)
J:		67 (7.8)
	J1	60 (7)
	>J1b	11 (1.3)
	>J1c	49 (5.7)
	J2	7 (0.8)
	>J2a	4 (0.5)
	>J2b	3 (0.4)
T:		91 (10.6)
	T1	37 (4.3)
	T2	54 (6.3)
N:		67 (7.8)
	N*	3 (0.4)
	N1	22 (2.6)
	>N1a	6 (0.7)
	>N1b1	6 (0.7)
	>I	10 (1.2)
	N2	24 (2.8)
	>W	24 (2.8)
	X2	18 (2.1)
M:		8 (0.9)
	C	2 (0.2)
	D4	3 (0.4)
	M*	1 (0.1)
	M1	1 (0.1)
	Z1	1 (0.1)
L:		3 (0.4)
	L0a	2 (0.2)
	L3f	1 (0.1)
Общо:		855

Почти половината от цялата извадка (42%) принадлежи към хаплогрупа H, което означава, че това е най-честата мтДНК хаплогрупа при българите. Сред подхаплогрупите на Хг H доминираща е H1 (8.5%). Останалите мтДНК хаплогрупи, произлизащи от макро-хаплогрупа R0, каквито са хаплогрупите HV0 и HV и подхаплогрупата R0a се срещат с по-ниски честоти (съответно 0.5%, 3.7% и 3.9%). Сред представителите на Хг U най-многобройни са тези класифицирани към U4 (3.9%), U5 (6.7%) и производната на U8- K (5.9%). От производните на хаплогрупния клъстер JT по-застъпени са T1 (4.3%), T2 (6.3%) и J1c (5.7%). Установените в българската популация N(xR) хаплогрупи: N1a, N1b1, I, W и X2 са редки (съответно 0.7%, 0.7%, 1.2%, 2.8% и 2.1%). Супер-хаплогрупа M е спорадично

представена от хаплогрупи C, D4, M*, M1 и Z1, а супер-хаплогрупа L има единични представители от L0a и L3f.

1.2. Вътрепопулационно сравнение

При вътрепопулационното сравнение се изследва зависимостта между хаплогрупната принадлежност и месторождението на майката чрез χ^2 тест. Той се провежда върху броя представители на мтДНК хаплогрупите H, HV, I, J, K, L, M, N, R, T, U, V, W и X в следните райони на страната: София-град и София-област, Североизточна, Северозападна, Югоизточна и Югозападна България (Фигура 3) (Таблица 3). Получените резултати са следните: $\chi 2$ =65.901 и p=0.093 с 52 степени на свобода. При ниво на значимост α=0.05 (p>α) не може да се отхвърли нулевата хипотеза за независимост между броя лица от районите на страната и броя лица принадлежащи към мтДНК хаплогрупите. Оттук следва, че разпределението на мтДНК хаплогрупите в България е равномерно.



Фигура 3. Райони на България включени във вътрепопулационното сравнение на честотата на мтДНК хаплогрупите.

Таблица 3. Брой представители на мтДНК хаплогрупите в различни райони на България.

	Брой лица					
Район / Хг	**София –град и област**	**Северо-източна България**	**Северо-западна България**	**Югоизточна България**	**Югозападна България**	**Общо**
H	71	24	52	98	95	340
HV	9	4	9	9	21	52
I	2	1	0	1	6	10
J	7	6	9	19	20	61
K	7	4	5	15	17	48
L	0	0	0	3	0	3
M	3	1	2	0	1	7
N	3	1	0	2	8	14
R	3	0	0	1	3	7
T	18	3	18	26	20	85
U (xK)	25	11	20	31	47	134
V	2	4	1	2	4	13
W	2	2	5	5	9	23
X	2	3	5	3	4	17
Общо	154	64	126	215	255	814

1.3. Междупопулационно сравнение на честотата на мтДНК хаплогрупи

Междупопулационното сравнение е проведено чрез анализ на главните компоненти (PCA). В него българската популация е сравнена с 41 популации от Западна Евразия и Северна Африка. PCA анализът е проведен върху 23 мтДНК хаплогрупи (H*, H5, HV0, HV, R0a, J, T1, T2, U1, U2e, U3, U4, U5a, U5b, U6, U7, U8, U*, K, N1, N2, X и M), а производните на супер-хаплогрупа L и хаплогрупите, които са различни или не произлизат от горните, не са включени във входните данни за анализа (Таблица 4). Наборът от 45 мтДНК хаплогрупи при българите е съкратен, поради нуждата от приравняване на нашите данни с тези от предишни проучвания проведени на по-ниско ниво на филогенетична резолюция. Напр. при българската популация хаплогрупите HV*, HV1 и HV2 са обединени в хаплогрупа HV, тъй като при други популации в PCA анализа Хг HV не е допълнително класифицирана. Понижаването на филогенетичната резолюция ни позволи към данните за българската популация да добавим резултатите за 141 българи, включени в предишно проучване на европейски мтДНК линии [Richards et al., 2002].

Таблица 4. Брой лица принадлежащи към мтДНК хаплогрупите и подхаплогрупите в извадките от 42 популации включени в PCA анализа.

Популация	Брой лица													
	Общо	H*	H5	HV0	HV	R0a	J	T1	T2	U1	U2e	U3	U4	U5a
Австрия [1]	99	43	3	1	1	1	10	2	8	0	1	1	4	8
Албания [1]	42	18	0	1	1	0	3	1	2	1	0	0	0	3
Англия [2]	335	148	13	11	1	0	46	7	22	0	3	2	7	13
Босна [2]	144	61	8	9	0	2	10	2	5	2	0	1	8	10
България [2,***]	**996**	**380**	**33**	**35**	**39**	**6**	**77**	**46**	**59**	**13**	**10**	**21**	**39**	**45**
Германия [1-6]	905	368	43	37	3	0	83	23	71	4	2	16	26	46
Гърция [1]	155	54	6	3	5	3	21	4	6	3	1	3	4	7
Египет [7-9]	413	15	2	13	14	7	29	9	17	2	1	14	5	1
Естония [1]	558	235	17	18	6	0	55	7	36	1	7	5	32	56
Ирландия [1]	300	126	5	17	4	1	33	4	17	0	4	3	4	11
Испания [2,10]	1029	437	33	76	8	3	84	17	51	1	2	12	19	33
Италия (Северна) [1]	346	131	33	18	8	1	21	6	26	3	5	4	10	7
Италия (Централна) [1]	1273	427	52	61	45	14	88	34	129	11	10	34	21	56
Италия (Южна) [1]	539	209	25	19	18	5	45	13	35	10	4	17	13	11
Кавказ [1]	2650	573	61	24	104	7	172	83	184	108	60	147	104	153
Кипър [11]	91	18	2	0	0	0	5	0	7	2	0	5	2	2
Крит [1]	202	79	7	2	6	0	11	6	22	4	1	16	1	4
Латвия[1]	299	113	20	9	7	0	19	5	23	0	9	5	28	21
Лемнос[1]	60	26	4	0	0	0	3	0	10	2	1	0	1	0
Норвегия[1]	556	250	17	21	1	0	58	4	44	1	0	8	17	35
Палестинци[1]	117	28	5	0	2	3	11	4	11	1	1	1	2	1
Поляци [1]	542	192	34	22	6	0	51	10	43	0	4	27	25	22
Португалия[1]	594	261	12	27	0	4	38	20	41	0	6	8	12	15
Родос [1]	42	14	1	0	0	0	3	0	6	0	0	5	0	1
Румъния [1]	94	26	7	6	1	2	8	8	5	0	0	3	4	7
Русия [2,12]	549	220	27	26	12	2	40	16	43	2	8	5	19	38
Сардиния [1]	370	165	4	25	4	0	45	6	34	4	1	5	1	7
Сицилия [1]	105	51	6	3	4	2	5	2	6	2	1	1	3	1
Скотландия	1199	474	36	42	1	0	168	31	95	6	8	18	23	60
Словакия[1]	129	49	9	3	2	0	19	2	9	0	2	3	2	9
Словения [1]	104	38	11	7	0	0	10	1	5	0	1	2	6	8
Страна на баските [1]	156	87	6	17	0	0	4	0	9	0	0	0	0	2
Турция [1]	340	99	10	2	17	0	32	11	18	11	4	19	5	4
Уелс [1]	92	43	7	3	2	0	14	2	2	0	0	0	0	3
Украйна [6]	110	44	5	2	1	0	14	2	7	1	2	1	6	7
Унгария [13,14]	533	136	21	25	4	7	46	8	21	4	4	1	9	15
Финландия [1,15]	312	113	8	19	0	0	19	5	13	1	2	0	5	18
Франция [1,16]	332	125	16	19	2	3	17	8	29	0	7	2	7	17
Хърватска [17]	96	36	7	5	3	0	3	2	4	1	4	2	2	8
Чехия [1]	83	28	6	5	3	0	7	3	8	0	1	1	1	7
Швейцария[1]	228	93	11	11	1	0	26	6	22	0	2	2	8	11
Швеция и Дания[1]	75	31	4	3	0	0	9	2	5	0	1	0	3	5

продължена

Популация	Брой лица									
	U5b	U6	U7	U8	U*	K	N1	N2	X	M
Австрия [1]	0	0	0	2	1	7	2	1	1	2
Албания [1]	3	0	0	0	0	3	1	0	0	1
Англия [2]	15	0	1	1	0	21	10	3	3	0
Босна [2]	7	0	0	0	0	6	5	2	2	2
България [2,***]	**25**	**0**	**7**	**4**	**3**	**59**	**27**	**25**	**20**	**11**
Германия [1-6]	35	0	2	2	3	63	24	21	10	2
Гърция [1]	4	0	2	0	1	7	7	2	7	4
Египет [7-9]	15	3	2	0	0	23	27	4	7	41
Естония [1]	24	0	0	9	1	15	13	15	5	1
Ирландия [1]	15	0	0	0	0	37	9	7	2	1
Испания [2,10]	23	16	0	3	34	66	14	24	18	8
Италия (Северна) [1]	4	0	0	1	0	32	12	6	17	1
Италия (Централна) [1]	34	4	14	10	1	84	35	25	39	25
Италия (Южна) [1]	7	7	6	2	0	39	21	9	14	7
Кавказ [1]	27	1	25	13	1	159	56	93	145	163
Кипър [11]	0	2	1	0	0	19	4	3	3	2
Крит [1]	6	0	1	0	0	15	3	0	8	2
Латвия[1]	6	0	0	0	0	7	13	12	1	1
Лемнос[1]	0	0	0	3	0	5	1	2	0	2
Норвегия[1]	31	0	0	2	0	31	13	9	2	6
Палестинци[1]	0	1	3	0	1	9	3	3	4	4
Поляци [1]	17	0	1	3	0	21	15	23	11	12
Португалия[1]	15	14	1	1	7	36	9	11	9	3
Родос [1]	0	0	2	1	0	3	2	1	1	0
Румъния [1]	5	0	0	0	0	3	0	6	2	0
Русия [2,12]	14	0	1	2	0	22	14	13	12	8
Сардиния [1]	30	1	1	0	0	17	4	1	5	8
Сицилия [1]	1	1	2	0	0	6	3	0	3	0
Скотландия	30	0	1	4	0	97	61	13	30	1
Словакия[1]	2	0	0	0	0	5	6	3	0	1
Словения [1]	3	0	0	0	0	4	2	5	1	0
Страна на баските [1]	17	0	1	1	0	6	0	0	2	0
Турция [1]	6	0	2	4	2	19	13	10	15	15
Уелс [1]	3	0	0	0	0	7	3	0	1	0
Украйна [6]	4	0	1	1	0	3	3	1	1	4
Унгария [13,14]	17	1	1	11	2	113	28	28	7	11
Финландия [1,15]	44	0	1	0	0	19	17	16	6	0
Франция [1,16]	12	2	1	3	0	35	8	5	4	6
Хърватска [17]	2	0	1	0	0	6	3	4	0	2
Чехия [1]	3	0	0	0	0	3	3	1	3	0
Швейцария[1]	5	1	0	1	0	12	3	4	1	0
Швеция и Дания[1]	0	0	0	0	0	8	1	1	0	0

* H(xH5), [1][Achilli et al., 2007], [2][Richards et al., 2000], [3]настоящото проучване, [4][Lutz et al., 1998], [5][Brandstatter et al., 2006], [6][Tetzlaff et al., 2007], [7]непубликувани данни, [8][Stevanovitch et al., 2004], [9][Coudray et al., 2009], [10][Saunier et al., 2009], [11][Alvarez-Iglesias et al., 2009], [12][Irwin et al., 2008],

[13][Irwin et al., 2010], [14][Brandstatter et al., 2007], [15][Brandstatter et al., 2008], [16][Hedman et al., 2007], [17][Richard et al., 2007], [18][Babalini et al., 2005]

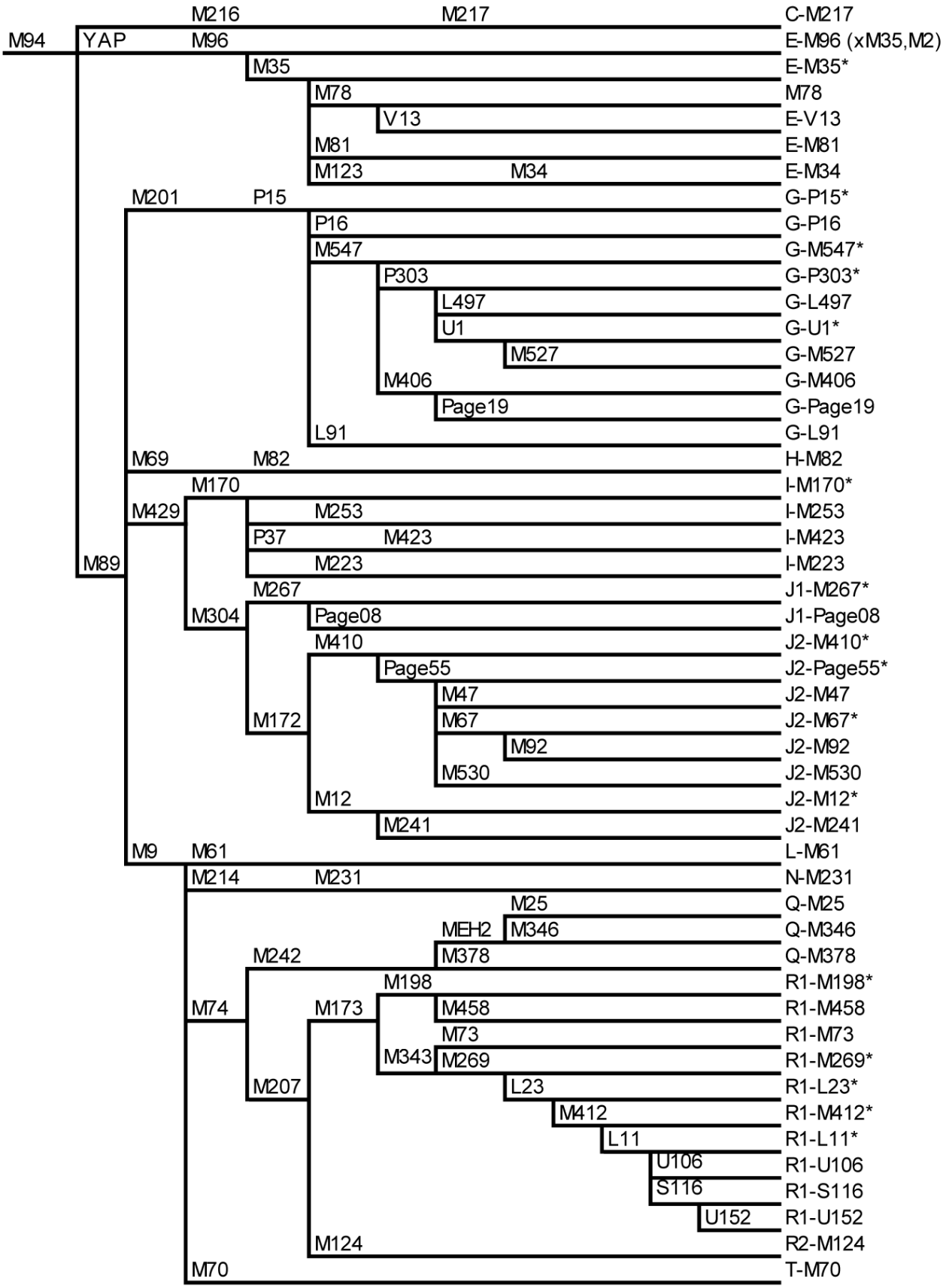
2. Определяне на Y-хромозомни хаплогрупи и хаплотипове при българите

2.1. Честота на Y-хромозомните хаплогрупи при българите

От 75 изследвани биалелни маркера на Y-хромозомата е установено наличието на 67, от които 49 са информативни т. е. определят филогенетично крайни хаплогрупи (Фигура 4). Генотипирани са мъже от цялата страна, които според мястото на раждане на бащата са класифицирани в деветте области от предишното административно деление на България в периода 1987-1998 г.: Бургас, Варна, Ловеч, Монтана, Пловдив, Разград, София-град, София-област и Хасково. Разделянето в тези области е по примера на най-мащабното антропологично проучване на българите [Йорданов, 2006].

Честотите на установените Y-хромозомни хаплогрупи в страната и в отделните области са представени в Таблица 5.

Резултатите показват че, най-честите Y-хромозомни хаплогрупи при българите са I-M423 (20.2%) и E-V13 (18.1%), следвани от хаплогрупа R-M17 (17.5%), от която 42.9% принадлежат към Хг R-M458. Хаплогрупа R-L23* се среща при 5.2% от българите, обхващайки почти половината от положителните по M269 Y-хромозоми. Следващи по честота са хаплогрупи I-M253 (4.3%), J-M241 (3.8%) и J-M530 (2.4%). Останалата част от филогенетично крайните хаплогрупи се срещат с честота по-ниска от 0.2%.



Фигура 4. Филогенетични връзки на Y-хромозомните хаплогрупи в българска популация. Маркерите M2 (от Хг E), M365, M367, M368, M390 (от Хг J), M120, M323 (от Хг Q) и M529 (от Хг R) са анализирани, но не са установени.

Таблица 5. Честота на Y-хромозомните хаплогрупи във всяка от областите според предишното административно деление на България (от 1987 до 1998 год) и в страната като цяло.

Брой (%) Хаплогрупа	Бургас	Варна	Ловеч	Монтана	Пловдив	Разград	София-град	София-област	Хасково	Неуточнена област	Общо
C-M217				1 (1.3)				2 (0.8)		1 (1.4)	**4 (0.5)**
E-M96(xM35,M2)	1 (2.2)	1 (6.7)			1 (0.6)						**3 (0.4)**
E-M35*			1 (1.6)								**1 (0.1)**
E-M78	1 (2.2)			3 (3.8)	1 (0.6)			7 (2.7)			**12 (1.5)**
E-V13	8 (17.8)	3 (20)	15 (24.2)	16 (20)	32 (20.1)	8 (38.1)	5 (8.5)	37 (14.4)	12 (29.3)	10 (14.5)	**146 (18.1)**
E-M81									1 (2.4)		**1 (0.1)**
E-M34	2 (4.4)			2 (2.5)	4 (2.5)		1 (1.7)	5 (1.9)		1 (1.4)	**15 (1.9)**
G-P15*							1 (1.7)			1 (1.4)	**2 (0.2)**
G-P16								1 (0.4)			**1 (0.1)**
G-M547*								1 (0.4)			**1 (0.1)**
G-P303*			2 (3.2)					3 (1.2)		1 (1.4)	**6 (0.7)**
G-L497	1 (2.2)			2 (2.5)	4 (2.5)	1 (4.8)		7 (2.7)			**15 (1.9)**
G-U1						2 (9.5)		2 (0.8)			**4 (0.5)**
G-M527				1 (1.3)							**1 (0.1)**
G-M406								1 (0.4)			**1 (0.1)**
G-PAGE19			1 (1.6)					1 (0.4)			**2 (0.2)**
G-L91	1 (2.2)				2 (1.3)		1 (1.7)	1 (0.4)		2 (2.9)	**7 (0.9)**
H-M82					1 (0.6)		2 (3.4)	1 (0.4)		1 (1.4)	**5 (0.6)**
I-M170*					2 (1.3)		1 (1.7)				**3 (0.4)**
I-M253		1 (6.7)	2 (3.2)	5 (6.3)	9 (5.7)		2 (3.4)	11 (4.3)	1 (2.4)	4 (5.8)	**35 (4.3)**
I-M423	7 (15.6)	4 (26.7)	9 (14.5)	15 (18.8)	32 (20.1)	2 (9.5)	15 (25.4)	59 (23)	2 (4.9)	18 (26.1)	**163 (20.2)**
I-M223		1 (6.7)	1 (1.6)		4 (2.5)		1 (1.7)	6 (2.3)	1 (2.4)		**14 (1.7)**
J1-M267*	2 (4.4)		1 (1.6)	2 (2.5)	1 (0.6)		1 (1.7)	4 (1.6)			**11 (1.4)**
J1-PAGE08	1 (2.2)		1 (1.6)	2 (2.5)	2 (1.3)		3 (5.1)	6 (2.3)	1 (2.4)		**16 (2)**
J2-M410*	1 (2.2)							2 (0.8)		1 (1.4)	**4 (0.5)**

продължена

Брой (%) Хаплогрупа	Бургас	Варна	Ловеч	Монтана	Пловдив	Разград	София-град	София-област	Хасково	Неуточнена област	Общо
J2-PAGE55*	2 (4.4)		4 (6.5)	2 (2.5)	3 (1.9)			2 (0.8)	1 (2.4)		**14 (1.7)**
J2-M47										1 (1.4)	**1 (0.1**
J2-M67					4 (2.5)			3 (1.2)		1 (1.4)	**8 (1)**
J2-M92							1 (1.7)	2 (0.8)	1 (2.4)	1 (1.4)	**5 (0.6)**
J2-M530			3 (4.8)	1 (1.3)	2 (1.3)	1 (4.8)	2 (3.4)	8 (3.1)	2 (4.9)		**19 (2.4)**
J2-M12*			1 (1.6)	1 (1.3)				1 (0.4)			**3 (0.4)**
J2-M241	5 (11.1)	1 (6.7)	3 (4.8)	2 (2.5)	6 (3.8)	1 (4.8)	2 (3.4)	10 (3.9)		1 (1.4)	**31 (3.8)**
L-M61			1 (1.6)							1 (1.4)	**2 (0.2)**
N-M231					1 (0.6)			1 (0.4)		2 (2.9)	**4 (0.5)**
Q-M25								1 (0.4)			**1 (0.1)**
Q-M346					1 (0.6)						**1 (0.1)**
Q-M378			1 (1.6)	1 (1.3)							**2 (0.2)**
R1-M198*	2 (4.4)	1 (6.7)	6 (9.7)	15 (18.8)	13 (8.2)		9 (15.3)	27 (10.5)	5 (12.2)	3 (4.3)	**81 (10)**
R1-M458	3 (6.7)	1 (6.7)	6 (9.7)	3 (3.8)	13 (8.2)	2 (9.5)	3 (5.1)	18 (7)	7 (17.1)	5 (7.2)	**61 (7.5)**
R1-M73			1 (1.6)		1 (0.6)						**2 (0.2)**
R1-M269*	1 (2.2)		1 (1.6)	1 (1.3)			2 (3.4)	2 (0.8)		1 (1.4)	**8 (1)**
R1-L23	4 (8.9)	2 (13.3)		2 (2.5)	8 (5)	3 (14.3)	3 (5.1)	11 (4.3)	4 (9.8)	5 (7.2)	**42 (5.2)**
R1-M412*	1 (2.2)				1 (0.6)						**2 (0.2)**
R1-L11*					1 (0.6)						**1 (0.1)**
R1-U106	2 (4.4)		1 (1.6)	1 (1.3)	2 (1.3)			4 (1.6)			**10 (1.2)**
R1-S116				1 (1.3)	1 (0.6)		2 (3.4)			2 (2.9)	**6 (0.7)**
R1-U152					5 (3.1)		1 (1.7)	7 (2.7)	2 (4.9)	2 (2.9)	**17 (2.1)**
R2-M124								1 (0.4)			**1 (0.1)**
T-M70			1 (1.6)	1 (1.3)	2 (1.3)	1 (4.8)	1 (1.7)	2 (0.8)	1 (2.4)	4 (5.8)	**13 (1.6)**
Общо	45	15	62	80	159	21	59	257	41	69	**808**

2.2. Структура на българския Y-хромозомен генофонд в страната

Географското подразделяне на Y-хромозомната изменчивост в България е оценено чрез йерархичен AMOVA анализ на:

-популации, населяващи административните области, определени според предишното разделение на страната (1987-1998 г.) (Фигура 5);

-популации, групирани в Източна, Централна и Западна България;

-популации, групирани според физико-географските райони на страната (Черноморското крайбрежие, югозападния район, Дунавската равнина и Тракия).



Фигура 5. Административни области, според предишното разделение на страната (1987-1998 г.) включени в AMOVA анализа на честотата на Y-хромозомните хаплогрупи.

Таблица 6. Резултати от проведения AMOVA анализ.

	Сравнение	**Компоненти на дисперсията**	**Процент вариация**
Без групиране	Между популациите	0.00173*	0.38
	Вътре в популациите	0.44967	99.62
Групиране в Източна, Централна и Западна България[а]	Между групите	0.00221*	0.49
	Между популациите в групите	0.00014	0.03
	Вътре в популациите	0.44967*	99.48
Обединяване във физико-географски райони [б]	Между групите	0.00088	0.19
	Между популациите в групите	0.00097	0.21
	Вътре в популациите	0.44967	99.59

* p≈0.05; [а] обединяване на области Бургас, Варна и Разград (Източна България), Монтана, София-

град и -област (Западна България) и Ловеч, Пловдив и Хасково (Централна България); [б] обединяване на Бургас и Варна (Черноморско крайбрежие), София-град и София-област (югозападен район), Ловеч, Монтана и Разград (Дунавска равнина) и Пловдив и Хасково (Тракия)

Генетичната вариация между популациите от административните области (1987-1998 г.) е малка (0.38%) и със стойност на критерия за проверка близка до нивото на значимост ($p \approx 0.05$). Получените резултати при групирането на областите във физико-географски райони не са статистически значими, а при групирането на областите в Източна, Централна и Западна България неочаквано се получава междугрупова изменчивост (0.49%) с малка статистическа значимост ($p \approx 0.05$). Тези наблюдения показват, че съвременният български Y-хромозомен генофонд е доста еднороден в страната и че Стара планина не е действала като бариера за генен поток.

2.3. Междупопулационно сравнение на честотата на Y-хромозомни хаплогрупи

Съвкупността от честотите на Y-хромозомните хаплогрупи е използвана за сравняване на българите с други популации чрез PCA анализ. Данните за сравнение са взети от литературни източници и са приравнени до възможно най-високо ниво на филогенетична резолюция (както при PCA анализа на честотите на мтДНК хаплогрупи). В зависимост от изследвания контингент са проведени три PCA анализа.

В първия PCA анализ са включени 42 евразийски и африкански популации, анализирани с филогенетична резолюция, сравнима с постигнатата в настоящото изследване. Литературните данни за 23 Y-хромозомни хаплогрупи в тези популации са представени в Таблица 7.

Таблица 7. Брой лица спадащи към Y-хромозомните хаплогрупи в извадките от 42 популации включени в PCA анализа.

Страна/район (етническа група)	Брой лица																					
	Общо	A-M91 и B-M60	C-RPS4C-Y	E-P2	E-M78	E-M81	E-M123	G-M201	H-M69	I-M170	J-M12	J-M410	J-M67	J-M267	L -M61	NO-M214	Q-M242	R-M124	R-M269	R-M17	R-M458	T-M70
Албания [1,2]	55				14			1		12	8		3	2					10	3	2	
Афганистан/Северен [3]	44							2				2			11		2	5		22		
Афганистан/Южен [3]	146		5					13	6		5	4		1	7		4	4	1	96		
Босна (босненци) [1,2]	84				11			3		42	3	2	5	2					3	2	11	
Босна (хървати) [1,2]	90				8			1		66	1	1							2	7	4	
България [4]	**808**		**4**	**4**	**158**	**1**	**15**	**40**	**5**	**215**	**34**	**38**	**13**	**27**	**2**	**4**	**4**	**1**	**86**	**78**	**60**	**13**
Грузия [1,2]	66				1		1	21		1		9	12	3	1			3	6	6	1	1
Гърция [1,2]	92		1		18		2	3		9	8	4	7	2	1				16	10	5	4
Гърция (област Македония) [1,2]	57				12		1	1	1	17	3	3	2	1					8	1	5	1
Египет (араби) [2,5]	147	4		7	26	12	10	13		1	1	12	4	31				1	3	4		12
Египет/Южен [6]	29				5	2	2			1		1		6								3
Етиопия (амхари) [7]	78	9		30	28		4				1			2								4
Етиопия (ороми) [7]	48	8		10	11		1						1	15								2
Йемен [8]	62			5			5	1				3	3	45								
Йордания- Мъртво море [2,9]	45				6		14					2	1	4								
Йордания-Аман [2,9]	101	1			9	4	5	6		5	3	7	6	41					6	2		
Ирак (Мадан араби) [10]	143			3	1		5	2						116	1		4	2	4			
Иран/Северен [3,11]	33		1					5			1	6	1	3	1	3	3	1	5	1		
Иран/Южен [3,11]	117	3		2	5		1	15	3		4	19	4	14	7	1	3	1	7	19		4
Кавказ (балкарци) [1,2]	38				1			11	1	1	1	6	2		2			3	3	5		
Македония (албанци) [1,2]	64				23		2	1		11	9		1	4					12		1	
Оман [2,5]	121	1	4	8	2		15	2	2		5	6	1	46	1	1			1	11		10

продължена

Страна/район (етническа група)	Брой лица																					
	Общо	A-M91 и B-M60	C-RPS4C-Y	E-P2	E-M78	E-M81	E-M123	G-M201	H-M69	I-M170	J-M12	J-M410	J-M67	J-M267	L -M61	NO-M214	Q-M242	R-M124	R-M269	R-M17	R-M458	T-M70
Пакистан/Северен [2,12]	86		12					7	10	1	2	2			10	3	4	4	2	14		
Пакистан/Южен [2,12]	91		1	1	2			4	2		2	13	2	6	13	1	2	9	3	29		
Полша [1,2]	99				4					18	1			1		1			18	23	33	
Румъния [13]	149				24			3		45	7	8	4	3		1	1	1	18	26	7	1
Русия/Белгород [2,14]	143		1		1			1		24		6				18			4	68	17	2
Русия/Ливни [2,14]	110				1					25		1				6	1		4	53	16	3
Русия/Псков[2,14]	132				4		1			22		1		1		32			5	48	16	
Русия/Унжа[2,14]	52				3					14		2		2		7			5	11	6	
Северна Африка (бербери)[15]	94				4	62	2	2				2	3	3					5			1
Словения [1,2]	75				2			2		23		1	1	1					16	27	2	
Судан (араби) [16]	102	3			17					4				46								
Сърбия и Босна (сърби)[1,2,17]	184			2	34			7		72	8	3	2	2		7			13	22	10	
Турция/Западна [2,18]	163	2	4	3	13		7	14	1	16	2	19	15	15	6	5	1	1	22	9	1	4
Турция/Източна [2,18]	208		1		10		13	26	1	4	6	35	8	20	13	9	6	2	25	17	1	5
Турция/Централна [2,18]	152		2		3	1	9	17	1	8	1	31	10	12	3	7	3	2	29	7	1	4
Украйна [1,2]	92				7		1	3	1	18	3	1	2			6	1		2	29	17	1
Унгария [1,2]	53				4		1	1		7			1						8	20	10	1
Хърватска[1,2]	89				7		1	1		39	3		1			2			11	16	8	
Хърватска/Осиек [1,2]	29				3			4		8	2		1							10	1	
Чехия[1,2]	75							3		8	4					2			25	18	13	1

[1][Battaglia et al., 2009], [2][Underhill et al., 2010], [3][Lacau et al., 2012], [4]настоящото проучване, [5][Luis et al., 2004], [6][Arredi et al., 2004], [7][Semino et al., 2002], [8][Cadenas et al., 2008], [9][Flores et al., 2005], [10][Al-Zahery et al., 2011], [11][Regueiro et al., 2006], [12][Sengupta et al., 2006], [13][Martinez-Cruz et al., 2012], [14][Balanovsky et al., 2008], [15][Alonso et al., 2005], [16][Hassan et al., 2008], [17][Regueiro et al., 2012], [18][Cinnioglu et al., 2004]

Във втория анализ на главните компоненти е изследвано родството по бащина линия между българите и 43 европейски популации по отношение на шест чести в Европа Y-хромозомни хаплогрупи (Таблица 8).

Таблица 8. Брой лица принадлежащи към Y-хромозомните хаплогрупи в извадки от 44 европейски популации включени в PCA анализа.

Страна/район (етническа група)	Брой лица						
	Общо	E-M35	G-M201	I-M170	J2-M172	R-M17	P-M45 (xM17)
Албания [1]	55	14	1	12	11	5	10
Босна (босненци) [1]	84	11	3	42	10	13	3
Босна (хървати) [1]	90	8	1	66	2	11	2
България [2]	**808**	**175**	**40**	**215**	**85**	**138**	**91**
Германия [3]	16	1		6		1	8
Гърция [1]	92	20	3	9	19	15	18
Гърция (област Македония) [1]	57	13	1	17	8	6	9
Испания (баски) [3]	45	1		3			40
Испания/Андалузия [3]	29	3		1	2		19
Испания/Каталония [3]	24	1	2	1	1		19
Италия/граница Тоскана-Лацио [4]	79	9	12	4	15	4	32
Италия/Калабрия [3]	37	5	3		8		13
Италия/Марке, част на Апенините [4]	27	4	4	1	6	2	9
Италия/остров Елба [4]	95	7	10	19	8	1	50
Италия/Северозападна Пулия [4]	46	2	6		8	3	24
Италия/Североизточен Лацио [4]	55	14	7	3	8	1	21
Италия/Североизточна [1]	67	2	8	7	9	7	28
Италия/Централен Марке [4]	59	3	4	5	21	1	22
Италия/Централна Тоскана [4]	41	4	3	2	7	2	19
Италия/Южен Лацио [4]	51	3	6	4	13	2	19
Италия/Южна Кампания [4]	84	15	8	9	14	2	24
Италия/Южна Пулия [4]	71	16	8	2	17	1	19
Македония (албанци) [1]	64	25	1	11	10	1	12
Полша[1]	99	4		18	1	56	18
Румъния [5]	149	24	3	45	19	33	20
Русия/Белгород [6]	143	1	1	24	6	85	4
Русия/Ливни [6]	110	1		25	1	69	5
Русия/Псков [6]	132	5		22	1	64	5
Русия/Унжа [6]	52	3		14	2	17	6
Словения [1]	75	2	2	23	2	29	16
Сърбия и Босна (сърби) [1,7]	184	34	7	72	13	22	23
Украйна [1]	92	8	3	18	6	46	3
Унгария [1]	53	5	1	7	1	30	8
Финландия/Остроботния [8]	40		2	8		3	1

продължена

Страна/район (етническа група)	Брой лица						
	Общо	E-M35	G-M201	I-M170	J2-M172	R-M17	P-M45 (xM17)
Франция [3]	23	2		4	3		13
Франция (баски) [3]	22			2	1		19
Холандия [3]	27	1		6		1	19
Хърватска [1]	89	8	1	39	4	24	11
Хърватска/Осиек [1]	29	3	4	8	3	11	
Чехия [1]	75		3	8	4	31	26
Швеция (саами) [8]	38			12		6	3
Швеция/Блекинге/ Криханстад [8]	41		2	19		7	10
Швеция/Вестерботен [8]	41	1	2	17		4	7
Швеция/Готланд [8]	40	1		20		5	8

[1][Battaglia et al., 2009], [2]настоящото проучване, [3][Semino et al., 2000], [4][Capelli et al., 2007], [5][Martinez-Cruz et al., 2012], [6][Balanovsky et al., 2008], [7][Regueiro et al., 2012], [8][Karlsson et al., 2006]

При третия PCA анализ е определена позицията на българския Y-хромозомен генофонд сред 26 азиатски популации въз основа на девет хаплогрупи, а именно тези за които има данни от предишни проучвания (Таблица 9).

Таблица 9. Брой лица класифицирани към Y-хромозомни хаплогрупи в българска и в извадки от 26 азиатски популации.

Страна/район (етническа група)	Брой лица									
	Общо	C-RPS4Y	DE-YAP x E-SRY4064	E-SRY4064	I-M170	J-M172	L-M20	O-M175	R-M173	R-M124
Алтайска Република/Югозападна (казахи) [1]	30	17				1		1	4	
Алтайска Република/Югоизточна (казахи) [1]	89	54				4		30		
Афганистан/Северен [2]	44					2	11		22	5
Афганистан/Южен [2]	146	5				9	7		97	4
България [3]	**808**	**4**		**178**	**215**	**85**	**2**		**224**	**1**
Индия/Западна (с индоевропейски език) [4]	59	2				9	5		17	9
Индия/Източна (с австро-азиатски език) [4]	64	1				7		34		7
Индия/Източна (с индоевропейски език) [4]	64	1				4		3	22	12
Индия/Северна (с индоевропейски език) [4]	80	2				10	3	1	32	13
Индия/Североизточна (с тибето-бирмански език) [4]	87	1						75	4	5
Индия/Централна (с дравидски език) [4]	50						2	15		
Индия/Централна (с индоевропейски език) [4]	21							6	4	
Индия/Южна (с дравидски език) [4]	303	6				36	39	33	40	22
Иран/Северен [5]	33	1				8	1	1	8	1

продължена

Страна/район (етническа група)	Брой лица									
	Общо	C-RPS4Y	DE-YAP x E-SRY4064	E-SRY4064	I-M170	J-M172	L-M20	O-M175	R-M173	R-M124
Иран/Южен [5]	117			8		27	7		29	1
Киргизстан [6]	13	1						4	8	
Китай (манджури) [7]	133	36	1					79		1
Китай (Ханци) [7]	1729	191	43	1	0	4	1	1291	22	4
Корея [7]	64	8						52		
Монголия [7]	34	15						11	3	
Непал [8]	143	8				13		30	51	25
Непал (Тхару) [9]	171	1	6	3		24	4	63	15	8
Русия/Казан (татари) [10]	38			1	7	4	1		10	
Таджикистан [10]	168	4		6		23	16	1	76	10
Тибет [8]	156	4	79			1		53	3	1
Туркменистан [10]	30					5			13	1
Узбекистан [10]	366	42	8	9	8	49	11	18	128	8
Япония [11]	263	14	103					143		

[1][Dulik et al., 2011], [2][Lacau et al., 2012], [3]настоящото проучване, [4][Sengupta et al., 2006], [5][Regueiro et al., 2006], [6][Karafet et al., 2001], [7][Zhong et al., 2011], [8][Gayden et al., 2007], [9][Fornarino et al., 2009], [10][Wells et al., 2001], [11][Nonaka et al., 2007]

2.4. Y-STR хаплотипове на чести Y-хромозомни хаплогрупи

Част от пробите (247), принадлежащи към честите Y-хромозомни хаплогрупи (E-V13, I-M423, J-M241, R-L23*, R-M458, R-U106 и R-U152) са анализирани допълнително за алели в 17 Y-STR локуса.

Нашите Y-STR данни за Хг E-V13, I-M423, J-M241, R-L23*, R-M458, R-U106 и R-U152 са сравнени с тези от предишни проучвания на други евразийски популации. Поради това, че при тях са изследвани различни Y-STR локуси, сравнението е извършено върху сбора от Y-STR локусите изследвани във всички сравнявани популации. Статистическите анализи са извършени върху набор от осем STR локуса, с изключение на Хг R-M458, за която сравнението е въз основа на седем локуса.

Датирането на хаплогрупи E-V13, I-M423, J-M241, M458, R-L23*, R-U106 и R-U152 при българската и други евразийски популации е представено в Таблица 10. При тези изчисления трябва да се имат предвид несигурностите в калибрирането на Y-STR молекулния часовник и повишаването на оценките на възрастта при наличие на редки несвойствени алели и/или различни епизоди на генен поток при формирането на популацията.

Филогенетичните връзки между Y-STR хаплотиповете в хаплогрупи E-V13, I-M423, J-M241, R-M458 и R-L23* са представени чрез MJ-мрежи (Фигура 6).

Таблица 10. Оценка на възрастта и съответните стандартни грешки на Хг E-V13, I-M423, J-M241, R-M458, R-L23*, R-U106 и R-U152 при българска и други евразийски популации.

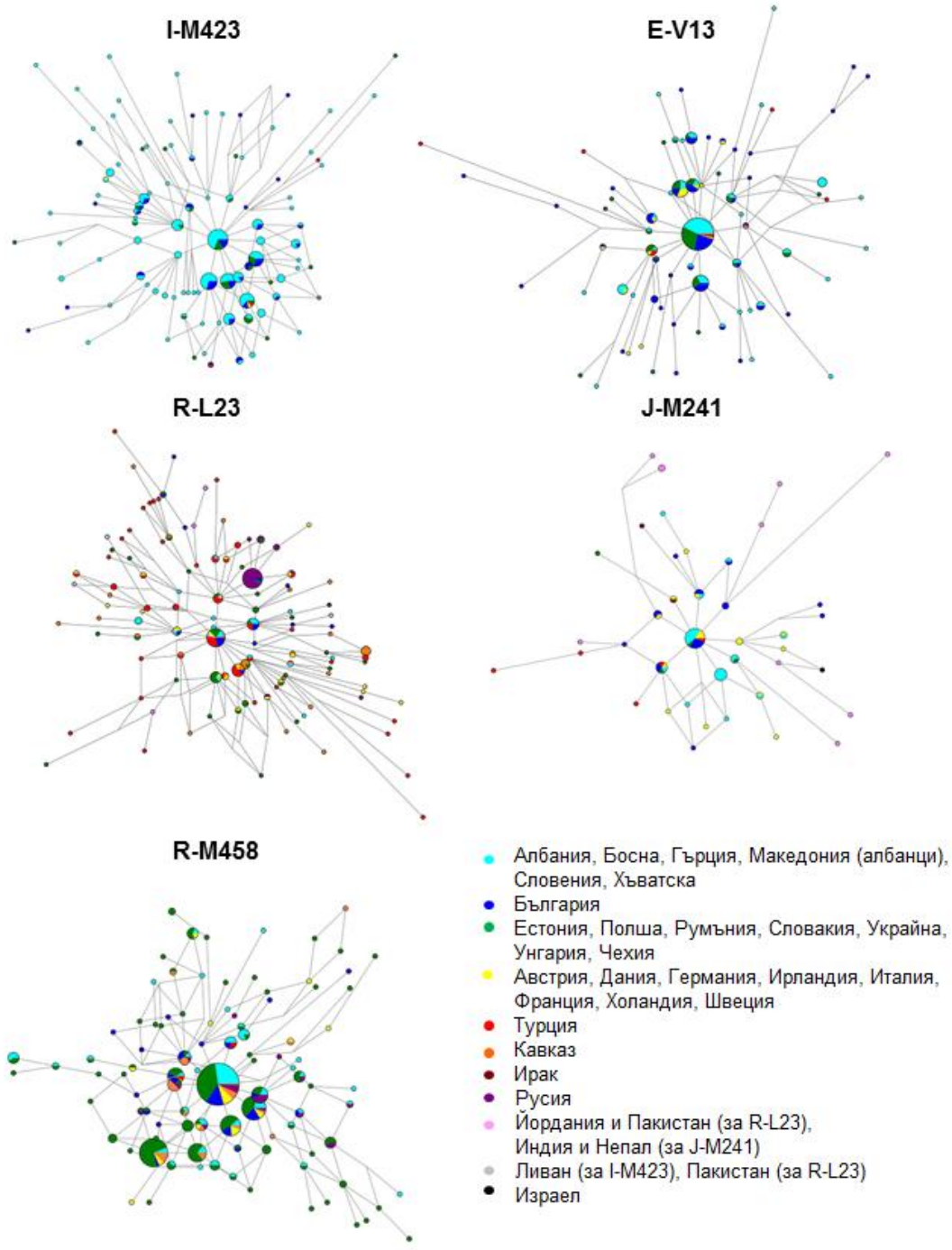
Оценките на възрастта са дадени в хиляди години преди (kya). Броят на пробите включени в анализа е даден в скоби.

Държава/район/ популация	Оценка на възрастта на хаплогрупа:						
	E-V13	**I-M423**	**J-M241**	**R-M458**	**R-L23***	**R-U106**	**R-U152**
Албания	5.0 ± 3 (12)[1]	11.9 ± 4 (8)[1]	4.0 ± 2 (8)[1]				
Англия						8.1 ± 3 (23)[2]	10.6 ± 1 (9)[2]
Босна (босненци)	3.9 ± 2 (11)[1]	7.5 ± 2 (36)[1]		4.7 ± 2 (11)[3]			
Босна (сърби и хървати)				1.9 ± 1 (8)[3]			
Босна (сърби)	6.2 ± 2 (16)[1]	10.0 ± 4 (27)[1]					
Босна (хървати)	15.1 ± 8 (8)[1]	8.7 ± 3 (43)[1]					
България						15.1 ± 8 (6)[4]	12.8 ± 4 (6)[4]
България/Западна	9.3 ± 3 (16)[4]	8.8 ± 3 (23)[4]	5.8 ± 3 (7)[4]	8.5 ± 4 (10)[4]	6.8 ± 3 (6)[4]		
България/Източна	8.4 ± 1 (18)[4]	7.9 ± 2 (12)[4]	3.6 ± 2 (5)[4]	12.4 ± 5 (5)[4]	16.8 ± 7 (6)[4]		
България/Централна	6.1 ± 2 (17)[4]	7.4 ± 3 (16)[4]	7.8 ± 3 (7)[4]	4.1 ± 1 (14)[4]	5.2 ± 3 (7)[4]		
Германия				10.0 ±3 (15)[5]		11.1 ±3 (61)[2]	11.2 ± 2 (33)[2]
Гърция			2.7 ± 1 (5)[1]				
Гърция/Атина	8.4 ± 3 (15)[1]	6.4 ± 3 (6)[1]		6.6 ± 3 (11)[3,5]	6.9 ± 2 (15)[2]		7.1 ± 3 (7)[2]
Гърция/Крит				8.6 ± 3 (7)[5]			
Гърция/област Македония	9.9 ± 5 (11)[1]	9.1 ± 2 (10)[1]		8.3 ± 4 (5)[3]			
Дания						9 ± 3 (19)[2]	
Естония				9.4 ± 5 (11)[5]		14.3 ± 6 (10)[2]	
Индия и Непал			20.1 ± 6 (10)[6]				
Ирак					14.0 ± 2 (12)[7]		
Иран					11.5 ± 2 (58)[7]		
Ирландия						9.4 ± 3 (6)[2]	
Италия					9.2 ± 2 (14)[2]	9.1 ± 2 (10)[2]	8.6 ± 1 (57)[2]

продължена

Държава/район/ популация	Оценка на възрастта на хаплогрупа:						
	E-V13	**I-M423**	**J-M241**	**R-M458**	**R-L23***	**R-U106**	**R-U152**
Италия/Калабрия	5.7 ± 3 (6)[1]						
Италия/Пулия	6.8 ± 3 (8)[3]		9.1 ± 4 (8)[3]				
Италия/Централна			4.5 ± 2 (5)[1]				
Кавказ				10.4 ± 3 (24)[3,5]	14.0 ± 3 (32)[2]		
Кавказ/Северозападен					14.0 ± 4 (6)[2]		
Кавказ/Североизточен					13.0 ± 3 (23)[2]		
Македония (албанци)	8.2 ± 3 (22)[1]	4.9 ± 2 (6)[1]	2.1 ± 1 (9)[1]				
Пакистан					14.5 ± 4 (5)[2]		
Полша		8.0 ± 3 (14)[1]		12.0 ± 4 (62)[3,5]	3.9 ± 1 (7)[2]	11.1 ± 2 (9)[2]	10.0 ± 2 (5)[2]
Румъния	4.3 ± 1 (22)[8]			10.4 ± 4 (7)[8]	14.3 ± 1 (12)[2]		7.9 ± 2 (6)[2]
Русия				10.1 ± 3 (12)[5]			
Русия/башкири					2.0 ± 1 (29)[2]		0 ± 0 (8)[2]
Русия/Вепса				3.4 ± 1 (6)[5]			
Русия/Карелия				3.0 ± 1 (12)[5]			
Словакия				10.9 ± 3 (56)[5]	7.2 ± 3 (10)[2]	12.0 ± 4 (11)[2]	9.7 ± 2 (7)[2]
Словения		14.5 ± 5 (15)[1]					10.6 ± 2 (6)[2]
Турция	10.6 ±3 (11)[1]		9.1 ± 2 (5)[1]		13.6 ± 2 (58)[2]		
Украйна	7.0 ± 2 (11)[1,8]	12.1 ± 5 (16)[1]		8.7 ± 3 (21)[3,5,8]			
Унгария	4.9 ± 3 (13)[1, 8]	5.4 ± 4 (5)[1]		10.7 ± 4 (25)[3,5,8]	9.7 ± 4 (7)[2]	12.4 ± 4 (6)[2]	
Франция						5.7 ± 1 (6)[2]	9.1 ± 2 (14)[2]
Холандия						8.1 ± 3 (28)[2]	
Хърватска	6.5 ± 4 (9)[1]	7.9 ± 3 (29)[1]		7.9 ± 2 (8)[3,5]			
Хърватска/Осиек		9.0 ± 3 (8)[1]		7.3 ± 2 (18)[3]			
Чехия				8.0 ± 3 (11)[3,5]			
Швейцария					7.7 ± 2 (10)[2]	10.0 ± 4 (19)[2]	8.6 ± 4 (29)[2]
Общо	**7.4 ± 2 (255)**	**9.1 ± 3 (274)**	**8.2 ± 2 (90)**	**10.4 ± 3 (384)**	**11.9 ± 1 (293)**	**11.0 ± 3 (234)**	**9.2 ± 1 (212)**

[1][Battaglia et al., 2009], [2][Myres et al., 2011], [3][Karachanak et al., 2013], [4]настоящото проучване, [5][Underhill et al., 2010], [6][Fornarino et al., 2009], [7]непубликувани данни, [8][Martinez-Cruz et al., 2012]



Фигура 6. MJ мрежи на Хг I-M423, E-V13, R-M458, R-L23* и J-M241.

За всяка хаплогрупа, кръговете и оцветените сектори са с размер, съответстващ на броя лица с еднакъв хаплотип. Най-малките кръгове представляват едно лице. Дължината на свързващите линии е пропорционална на броя на мутационните събития, различаващи два хаплотипа.

3. Генетична структура на европейски популации въз основа на SNPs в целия геном

Генетичната структура на европейските популации е изследвана чрез анализ на повече от 270 000 SNPs с платформа Illumina Infinium при 3 112 лица от 16 европейски страни: Австрия, България, Германия, Естония, Италия, Испания, Латвия, Литва, Полша, Русия, Унгария, Финландия, Франция, Чехия, Швеция и Швейцария. Българската извадка се състои от 47 наши ДНК проби.

За да се създаде карта на генетичната структура на Европа е проведен PCA анализ на данните за 266 356 от изследваните SNPs. Анализът е проведен върху 1 564 от лицата изследвани в проучването, сред които са и българите и 203 лица анализирани в HapMap проекта (жители на Юта с произход от Северна и Западна Европа, Хан китайци от Пекин, японци от Токио и Йоруба от Нигерия).

ОБСЪЖДАНЕ

Обсъждане на генетичната изменчивост на българите от гледна точка на мтДНК

Последователностите на контролния регион на мтДНК получени при секвенирането са групирани в 557 хаплотипа. Те са класифицирани в 45 мтДНК хаплогрупи и подхаплогрупи.

Българският митохондриален генофонд е почти изцяло съставен от хаплогрупи типични за западноевразийските популации. Представителите на източноазиатски (M, C, D и Z) и африкански (L0, L3 и M1) мтДНК хаплогрупи са единични, като общата им честота възлиза на 1.3%.

Супер-хаплогрупа R0

Генофондът на българите обхваща почти целия спектър на мтДНК хаплогрупи, произлизащи от основния ствол на супер-хаплогрупа R0, а именно: R0a, HV, HV0, HV0a, V и H. Подобно на почти всички европейски и близкоизточни популации, най-срещаната мтДНК хаплогрупа при българите е *H* (41.9%). Сред нейните подразделения най-честа е H1 (8.5%), за която се предполага, че е участвала в повторното заселване на Европа след последния ледников период (~22-19 kya) и след застудяването през младия дриас (~11.5 kya) [Roostalu et al., 2007] (Фигура 8). Някои от установените подхаплогрупи на Хг H, като H5 и H6, са по-чести и типични за Близкия Изток, Кавказ

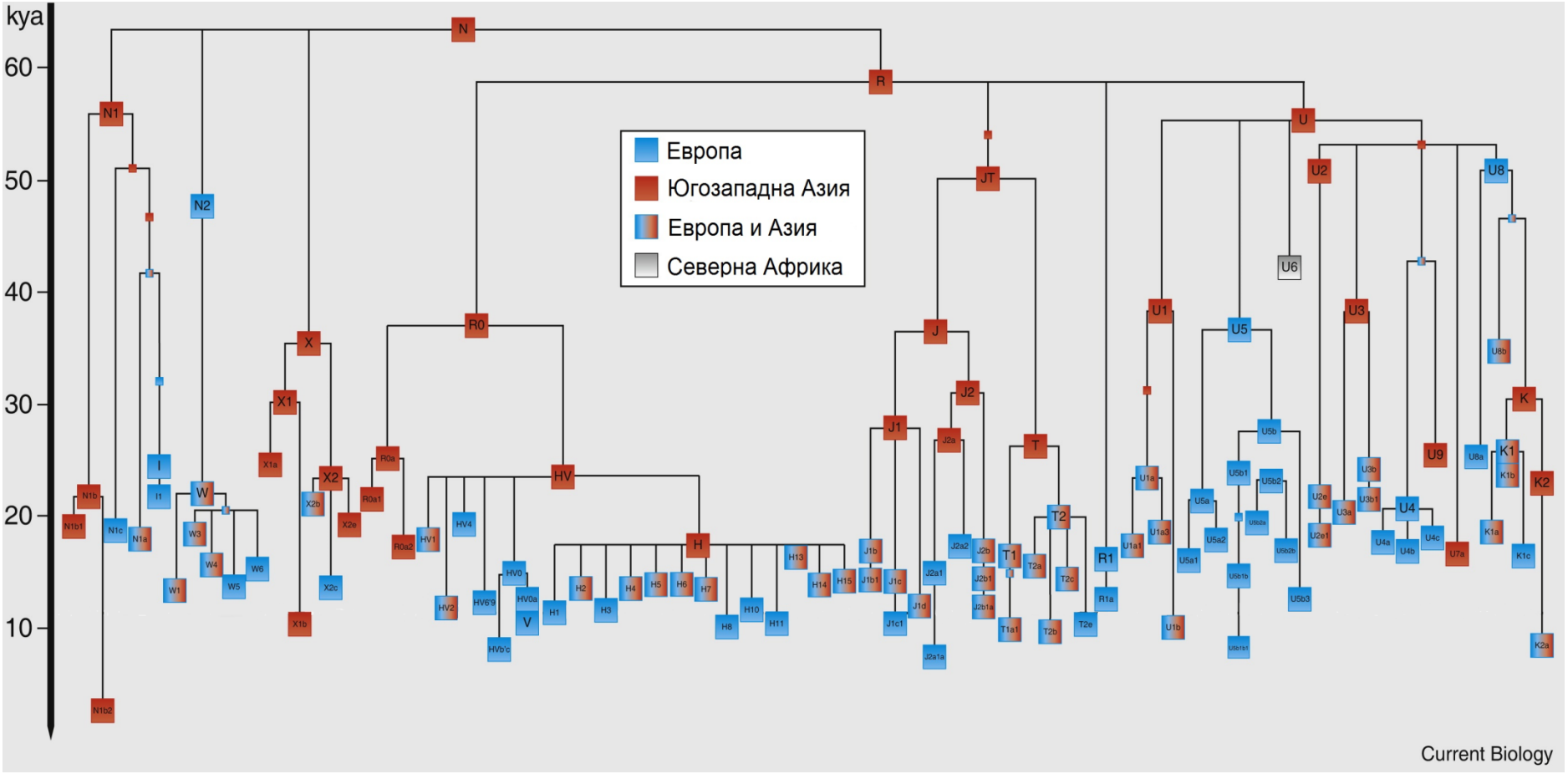
и Централна Азия. Тяхното присъствие сред българите (съответно 3.4% и 1.6%) не е изненадващо, имайки предвид историята на България и позицията й на предполагаемия път на азиатските мтДНК линии към Европа. *Хаплогрупа HV0* (3.7%), която е сестрински клон на Хг H, е рядко разпространена в Европа, достигаща по-висока честота в Средиземноморието (особено в Северна Африка) [Garcia et al., 2011]. Нейната производна V има най-висока честота на Иберийския полуостров [Torroni et al., 2001]. При българите, Хг V представлява 1.5% от общата популация, а *HV* клъстерът (без H/HV0) се наблюдава с честота 3.8%. Последният се среща с ниска честота на разпространение от Източното Средиземноморие до Южна Европа и достига максимална такава в Близкия Изток. Хг *R0a* се характеризира с почти същото географско разпространение, но има най-висока честота на Арабския полуостров. Тази мтДНК линия, произлизаща от основата на супер-хаплогрупа R0 при българите се среща с незначителна честота (0.5%).

Хаплогрупа U (включително K)

В съвременната българска популация хаплогрупа U е представена от Хг U1, U2, U3, U4, U5, U7, U8 и K. В тази мозайка от мтДНК хаплогрупи на различна възраст и с различно миграционно поведение, най-често срещаща се е хаплогрупа *U5* (6.7%), за която се предполага, че е произлязла от основата на клъстер U при първоначалното заселване на Европа, 45 kya [Soares et al., 2009]. Пробите принадлежащи към тази хаплогрупа са подкласифицирани към U5a и U5b (съответно 4.4% и 2.3%). Хаплогрупите *U1, U2* и *U4* са по-редки от Хг U5 (съответно 1.4%, 1.4% и 3.9%). Хаплогрупа *U8*, която е широко разпространена в Европа с ниски честоти е представена с подхаплогрупите U8a (0.2%) и U8b (0.2%). Подхаплогрупа U8a е по-честа в Западна Европа, а U8b се среща и извън Европа - в Кавказ, Иран, Близкия Изток и Северна Африка. Нейната подхаплогрупа *K* (5.9%) е по-честа и по-широко разпространена, покривайки ареала на U8a и U8b и достига дори до Индия [Gonzalez et al., 2006]. Болшинството от Хг K в България принадлежи към подхаплогрупа K1 (5.3%) и много по-рядко към подхаплогрупа K2 (0.6%). Хаплогрупите *U3* и *U7* (съответно 1.9% и 0.5%) са типични за Близкия Изток и за популациите от околните райони.

Хаплогрупа JT

Хаплогрупа *J* обхваща 7.9% от генетичната изменчивост на българите. Тя е представена от подхаплогрупите J1 (7%) и J2 (0.9%). По същия начин, хаплогрупа *T* е представена от подхаплогрупите T1 (4.3%) и T2 (6.3%).



Фигура 8. Филогенетично дърво на човешките мтДНК хаплогрупи, които са често наблюдавани при популациите от Европа, Югозападна Азия и Северна Африка. Времевата скала на възникване на хаплогрупите отразява времето преди хиляди години (kya). По-важни периоди в праисторията: горен палеолит 50-11.5 kya; последен ледников период 22-19 kya; млад дриас 11.5–11 kya и неолит ~10 kya. Хаплогрупите са обозначени с различен цвят в зависимост от региона от който произхождат (адаптирано от [Soares et al., 2010]).

Супер-хаплогрупа N

От хаплогрупите, които произлизат директно от основния ствол N, Хг *N1* е рядка в България (0.7%). Тя е еднакво представена от подхаплогрупи N1a и N1b1, които са с югозападноазиатски произход и се наблюдават с ниска честота сред европейски и близкоизточни популации [Derenko et al., 2007]. Подхаплогрупа N1b1 се среща и при северноафрикански популации [Palanichamy et al., 2010]. Останалата част от преките производни на супер-хаплогрупа N, а именно: *I, X2* и *W* се срещат с ниска честота (съответно 1.2%, 2.1% и 2.8%).

Супер-хаплогрупи M и L

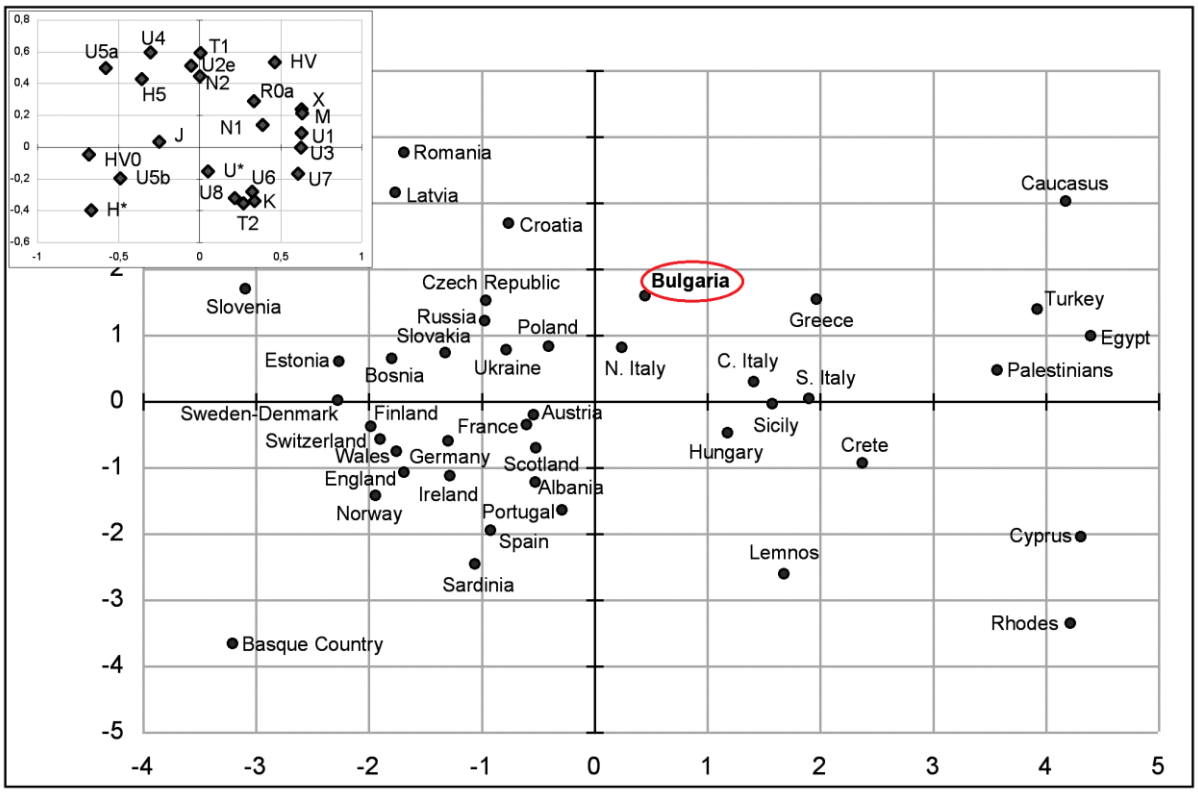
Специфичната за Азия супер-хаплогрупа *M* е установена с много ниска честота (0.9%). Тя е представена от източноазиатски (M*, C, D4 и Z1) и африкански (M1a) линии. Африканските хаплогрупи *L0a* и *L3f* са установени съответно в две и едно изследвани лица. Тези единични наблюдения не са изненадващи, защото ниски честоти на източноазиатски и африкански мтДНК хаплогрупи са описани и при други европейски популации.

От направения вътрепопулационен анализ се вижда, че няма статистически значима разлика в честотата на мтДНК хаплогрупите между София и София-област, Североизточна, Северозападна, Югоизточна и Югозападна България. Това показва, че българската мтДНК изменчивост е значително хомогенна в районите на страната.

Хомогенната структура на българския мтДНК генофонд позволи междупопулационните сравнения да се извършват въз основа на цялата българска извадка.

Сравнителен анализ на честотата на мтДНК хаплогрупи между различни популации

Родствените връзки по майчина линия между българска и други популации са определени чрез PCA анализ на честотата на мтДНК хаплогрупите. При този анализ, популациите са позиционирани в двуизмерна графика на първите два главни компонента (PCs) (Фигура 9), които заедно обясняват 32% от общата генетична информация.



Фигура 9. PCA анализ: графика на първите два главни компонента. В горния ляв панел е дадена графиката показваща приноса на всяка мтДНК хаплогрупа към първия и втория PC.

Първият главен компонент (ординатната ос) отразява 20% от дисперсията и разкрива географски градиент от запад към изток. Той се дължи на разпределението на Хг H* и HV0 (по-разпространени сред западноевропейските популации) от една страна и на хаплогрупите U1, U3, X, HV и U7 (най-чести сред кавказките и близкоизточните популации) от друга страна. Вторият главен компонент (абсцисната ос) обхваща 12% от дисперсията. С малки изключения той разделя европейските популации по географска дължина. Това се дължи на по-високи честоти на хаплогрупите U4 и T1 при източноевропейските и на хаплогрупите U5b и U8 при западноевропейските популации.

България се намира в централната част на графиката, което най-вероятно отразява смесването на популации с различен произход при формирането на съвременния български народ. Средиземноморският принос, който определя позицията на България по-далеч от останалите източноевропейски страни може да се припише на местните балкански популации-особено на траките, чиято велика цивилизация е достигнала своя апогей през VI век пр. н.е. От друга страна, източният принос може да се дължи на древните българи, които са се слели със славяните при мигрирането им към Балканите през ранното средновековие.

Обсъждане на генетичната изменчивост на българите от гледна точка на Y-хромозомата

Честотата на установените хаплогрупи показва, че българският Y-хромозомен генофонд е почти изцяло обхванат от западноевразийски хаплогрупи. Генетичният принос от Централна Азия (Хг C-M217), Северна Евразия (Хг N-M231) и Югозападна Азия (производни на Хг Q-M242, Хг LM61 и Хг R-M124) е почти незначителен (общата честота на тези хаплогрупи е 1.5%) подобно на други европейски популации.

Най-разпространените Y-хромозомни хаплогрупи при българите са I-M423 (20.2%) и E-V13 (18.1%). Те представляват автохтонни и почти ендемични подразделения на съответните хаплогрупи I-P37 и E-M78 в Югоизточна Европа [Battaglia et al., 2009; Cruciani et al., 2007]. Трета по честота е широко разпространената евразийска хаплогрупа R-M17, която е установена при 17.5% от българите, като 42.9% от тях принадлежат към типичната за Европа Хг R-M458. Хаплогрупа R-L23* (източното разклонение на западноевразийската хаплогрупа R-M269) е с честота 5.2% от българите, представлявайки почти половината от M269 произлезлите Y-хромозоми. Следваща по честота е хаплогрупа I-M253 (4.3%), която характеризира болшинството от индивидите принадлежащи към Хг I-M170 в Северна Европа. С по-ниска честота са две подразклонения на хаплогрупа J-M172, а именно: J-M241 и J-M530, съответно 3.8% и 2.4%. Останалата част от филогенетично крайните хаплогрупи притежават честоти по-ниски от 0.2%.

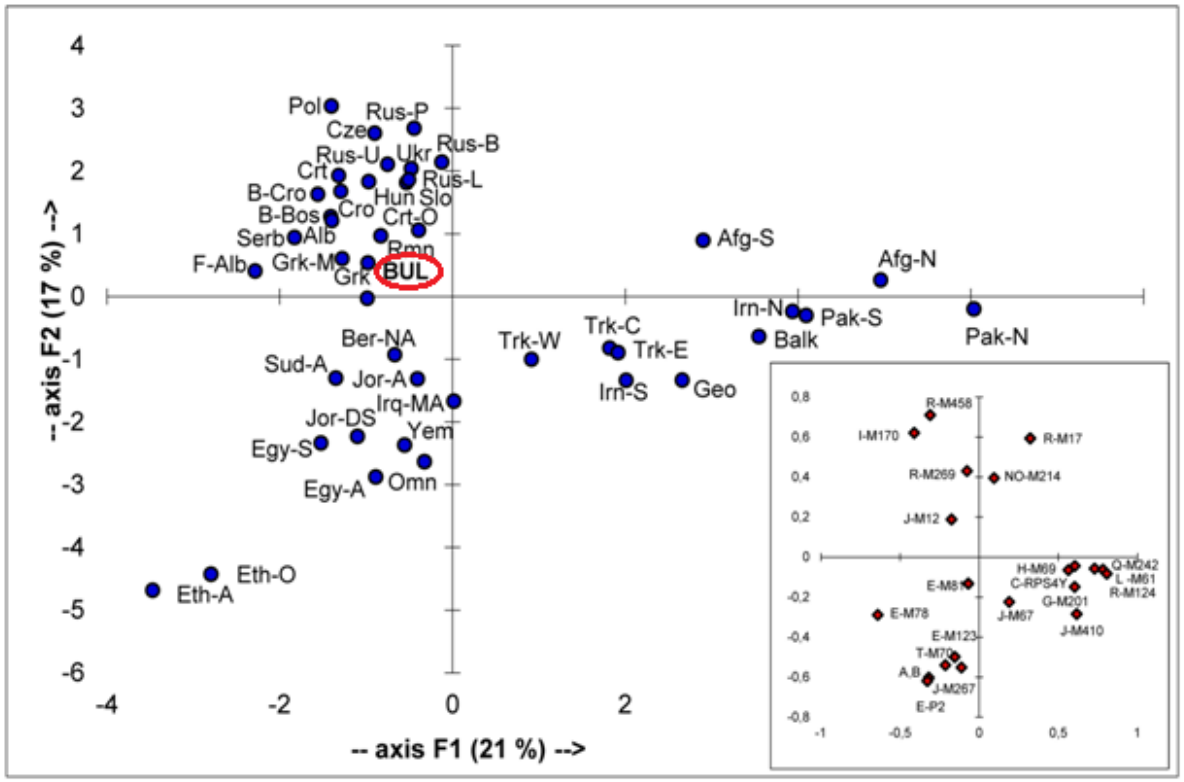
Резултатите от AMOVA анализа проведен на различни нива (върху предишните административни области и при тяхното групиране според географски и физико-географски критерии) показват, че българският Y-хромозомен генофонд е доста еднороден в страната и че Стара планина най-вероятно не е действала като силна бариера за генен поток.

Сравнителен анализ на честотата на Y-хромозомни хаплогрупи между различни популации

Детайлно описаният Y-хромозомен профил на българите може да се използва за определяне на тяхното място в по-широк контекст. Един от подходите за целта е сравняване с различни популации чрез PCA анализ въз основа на хаплогрупните честоти.

Междупопулационният анализ е проведен чрез три PCA анализа.

В първия PCA анализ са включени евразийски и африкански популации, анализирани на еднакво с нашето проучване високо ниво на филогенетична резолюция. Получената графика на първите два главни компонента, които заедно представляват 38% от дисперсията, е показана на Фигура 10.



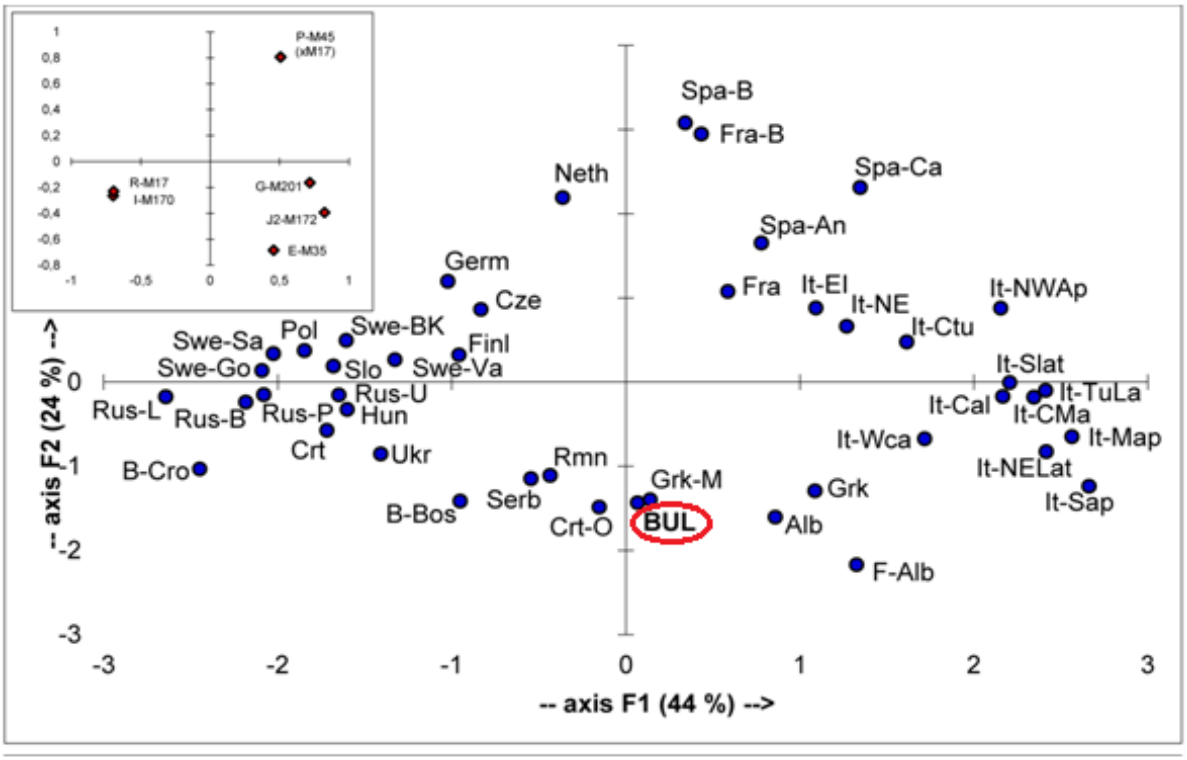
Фигура 10. Графика на първите два главни компонента при PCA анализ на африкански и евразийски популации, анализирани на най-високо ниво на филогенетична резолюция.

Страните, районите или популациите в тях (дадени в скоби) включени в анализа са дадени със следните съкращения: Afg-N: Северен Афганистан, Afg-S: Южен Афганистан, Alb: Албания, B-Bos: Босна (босненци), B-Cro: Босна (хървати), Balk: Кавказ (балкарци), Ber-NA: Северна Африка (бербери), BUL: България, Crt: Хърватска, Crt-O: Хърватска-Осиек, Cze: Чехия, Egy-A: Египет (араби), Egy-S: Южен Египет, Eth-A: Етиопия (амхари), Eth-O: Етиопия (ороми), F-Alb: Македония (албанци), Geo: Грузия, Grk: Гърция, Grk-M: Гърция (гърци от областта Македония в Гърция), Hun: Унгария, Irn-N: Северен Иран, Irn-S: Южен Иран, Irq-MA: Ирак (Мадан араби), Jor-A: Йордания-Аман, Jor-DS: Йордания- район на Мъртво море, Omn: Оман, Pak-N: Северен Пакистан, Pak-S: Южен Пакистан, Pol: Полша, Rmn: Румъния, Rus-B: Русия-Белгород, Rus-L: Русия-Ливни, Rus-P: Русия-Псков, Rus-U: Русия (Унжа), Serb: Сърбия и Босна (сърби), Slo: Словения, Sud-A: Судан (араби), Trk-C: Централна Турция , Trk-E: Източна Турция , Trk-W: Западна Турция, Ukr: Украйна и Yem: Йемен. На осите са посочени процентите, които показват частта от дисперсията, съдържаща се в главните компоненти. Вмъкнатата графика илюстрира приноса на хаплогрупите към главните компоненти.

В PCA графиката се разграничават четири основни клъстера от: източноафрикански, европейски, северноафрикански и близкоизточни арабски популации и югозападноазиатски клъстер формиран от турци, иранци, кавказци и

пакистанци. Първият главен компонент (axis F1, 21% от дисперсията) подрежда популациите по географска дължина, което в отрицателната част на оста основно се дължи на хаплогрупи I-M170 (почти ограничена само за Европа) и E-M78 (често срещана в Източна Африка и Югоизточна Европа), а в положителната - на хаплогрупи L-M61 и R-M124 (чести в Южна Азия). Вторият главен компонент (axis F2, 17% от дисперсията) разделя популациите по географска ширина, като хаплогрупи R-M17 и R-M458 са най-чести сред европейците; хаплогрупи A-M91, B-M60 и E-P2 се срещат почти единствено при източноафрикански групи и Хг E-M123 и J-M267 са най-чести в Северна Африка и Южния Среден Изток. Като цяло, европейците са разделени от турците, популациите от Кавказ, иранците и пакистанците по първия главен компонент и от арабските и източноафриканските популации по втория. Българите са локализирани в Европейския клъстер, в непосредствена близост до популацията на Северна Гърция, но относително далеч от югоизточните съседи – турците.

Анализът на главните компоненти чрез който се изследва родството по бащина линия между българите и европейските популации се извършва въз основа на най-честите хаплогрупи в Европа, които са с ниска филогенетична резолюция. PCA графиката на първите два компонента, които обясняват 68% от дисперсията е представена на Фигура 11.



Фигура 11. PCA графика на първите два компонента за анализ на българска популация в европейски контекст при по-ниско ниво на филогенетична резолюция.

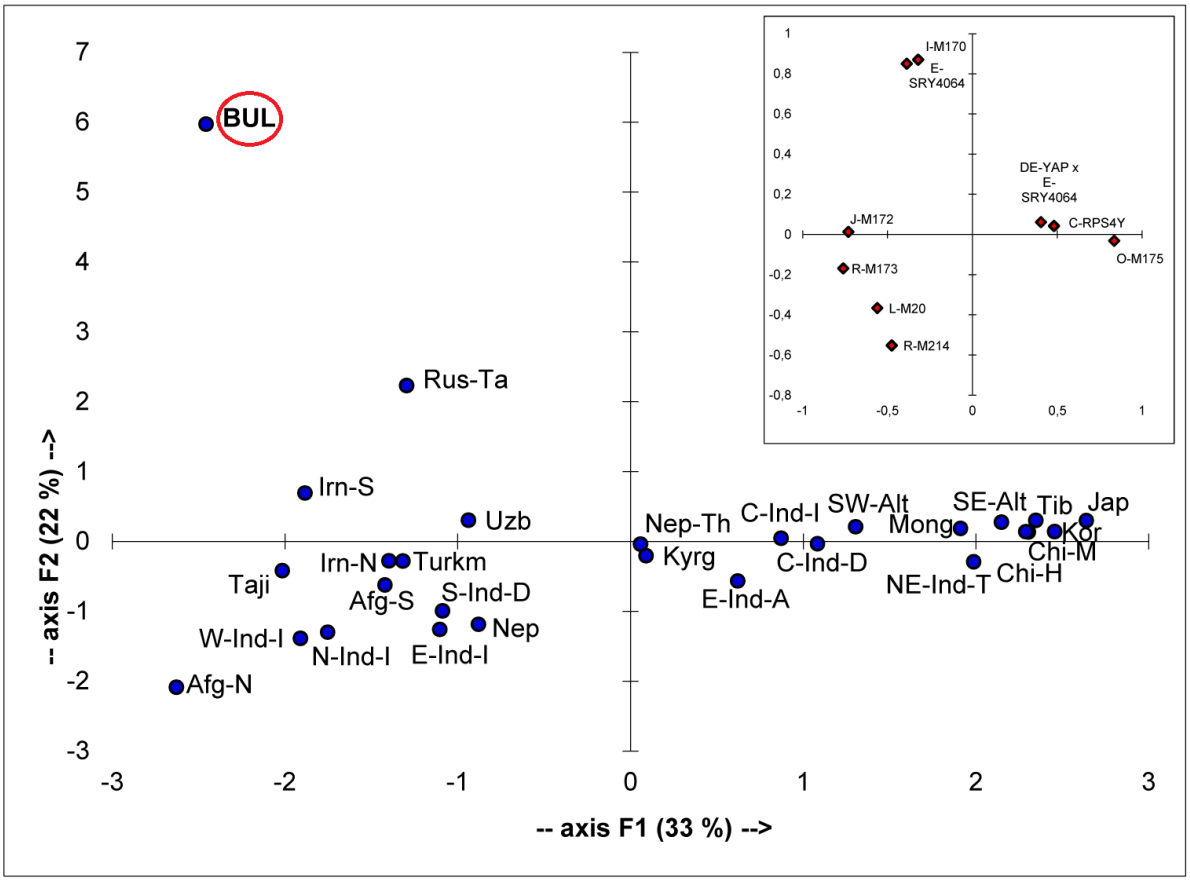
Съкращенията на страните, районите или популациите в тях (в скоби) са както следва: Alb: Албания, B-Bos: Босна (босненци), B-Cro: Босна (хървати), BUL: България, Crt: Хърватска, Crt-O: Хърватска-Осиек, Cze: Чехия, F-Alb: Македония (албанци), Finl: Финландия, Fra: Франция, Fra-B: Франция (баски), Germ: Германия, Grk: Гърция, Grk-M: Гърция (гърци от областта Македония в Гърция), Hun: Унгария, It-Cal: Италия-Калабрия, It-CMa: Италия-централна част на район Марке, It-CTu: Италия-Централна Тоскана, It-Elb: Италия-остров Елба, It-MAp: Италия-Марке, частта на Апенините, It-NEСевероизточна Италия, It-NELat: Италия-Североизточен Лацио, It-NWAp: Италия- Северозападна Пулия, It-SAp: Италия-Южна Пулия, It-SLat: Италия-Южен Лацио, It-TuLa: Италия-границата между Тоскана и Лацио, It-WCa: Италия-Южна Кампания, Neth: Холандия, Pol: Полша, Rmn: Румъния, Rus-B: Русия-Белгород, Rus-L: Русия-Ливни, Rus-P: Русия-Псков, Rus-U: Русия-Унжа, Serb: Сърбия и Босна (сърби), Slo: Словения, Spa-An: Испания-Андалузия, Spa-B: Испания (баски), Spa-Ca: Испания-Каталония, Swe-BK: Швеция-Блекинге,Криханстад, Swe-Go: Швеция-Готланд, Swe-Sa: Швеция (саами), Swe-Va: Швеция-Вестерботен и Ukr: Украйна. С числата в скобите е посочена частта от общата генетична информация запазена от главните компоненти. На по-малката графика е представен приносът на хаплогрупите към първия и втория главен компонент.

Разпределението на популациите по първия главен компонент (axis F1, обхващащ 44% от дисперсията) се определя от преобладаването на Хг I-M170 на Скандинавския полуостров и източното адриатическо крайбрежие и на Хг R-M17 в Северна-Централна и Североизточна Европа, както и на нарастващата честота на Хг J-M172 към Италия и Южните Балкани. Разделянето по втория главен компонент (axis F2, 24% от вариацията) основно се дължи на хаплогрупи E-M35 (най-разпространена в Южните Балкани) и на P-M45(xR-M17) (най-честа сред западноевропейските популации). Според предишни проучвания Хг P-M45(xR-M17) е предимно представена от Хг R-M269 [Alonso et al., 2005; López-Parra et al., 2009; Myres et al., 2011]. Този PCA анализ потвърждава положението на българите в близост до северните гърци и в съседство на славянските популации от Западните Балкани, за разлика от останалите изследвани славянски популации.

В третия PCA анализ на честотата на Y-хромозомните хаплогрупи, българите са сравнени с азиатски популации. Получената графика на първите два главни компонента (axis F1 и F2, обхващащи 55% от общата генетична информация) е представена на Фигура 12.

Популациите в графиката показват разпределение по географска ширина. Това основно се дължи на първия главен компонент (33% от дисперсията) или по-точно на преобладаването на източноазиатската Хг O-M175 и централноазиатската Хг C-RPS4Y на изток и на нарастващата честота на хаплогрупите J-M172 и R-M173 на запад. Вторият главен компонент (22% от дисперсията) основно е повлиян от хаплогрупи I-M170 и E-SRY4064, които ясно разделят българите от останалите популации. Като цяло

българите са отдалечени от алтайските популации и популациите живеещи на север от региона на Памир, и от казанските татари и иранци, макар и в по-малка степен.



Фигура 12. Графика на първите два главни компонента при PCA анализ на българите и някои азиатски популации въз основа на честотите на филогенетично уеднаквени Y-хромозомни хаплогрупи.

Изследваните страни, райони или популациите от тях (дадени в скоби) са обозначени със следните съкращения: Afg-N: Северен Афганистан, Afg-S: Южен Афганистан, BUL: България, C-Ind-D: Централна Индия (с дравидски език), C-Ind-I: Централна Индия (с индоевропейски език), Chi-H: Китай (Ханци), Chi-M: Китай (манджури), E-Ind-A: Източна Индия (с австро-азиатски език), E-ind-I: Източна Индия (с индоевропейски език), Irn-N: Северен Иран, Irn-S: Южен Иран, Jap: Япония, Kor: Корея, Kyr: Киргизстан, Mong: Монголия, N-Ind-I: Северна Индия (с индоевропейски език), NE-Ind-T: Североизточна Индия (с тибето-бирмански език), Nep: Непал, Nep-Th: Непал (Тхару), Rus-Ta: Русия-Казан (татари), S-Ind-D: Южна Индия (с дравидски език), SE-Alt: Югоизточна Алтайска Република (казахи), SW-Alt: Югозападна Алтайска Република (казахи), Taj: Таджикистан, Tib: Тибет, Turkm: Туркменистан, Uzb: Узбекистан и W-Ind-I: Западна Индия (с индоевропейски език). Частта от дисперсията запазена от компонентите е дадена в скобите. На вмъкнатата графика е показан приносът на хаплогрупите към главните компоненти.

Популациите в графиката показват разпределение по географска ширина. Това основно се дължи на първия главен компонент (33% от дисперсията) или по-точно на

преобладаването на източноазиатската Хг O-M175 и централноазиатската Хг C-RPS4Y на изток и на нарастващата честота на хаплогрупите J-M172 и R-M173 на запад. Вторият главен компонент (22% от дисперсията) основно е повлиян от хаплогрупи I-M170 и E-SRY4064, които ясно разделят българите от останалите популации. Като цяло българите са отдалечени от алтайските популации и популациите живеещи на север от региона на Памир, и от казанските татари и иранци, макар и в по-малка степен.

Време и място на произход на Y-хромозомните хаплогрупи според Y-STR изменчивостта

Оценката на възрастта (Таблица 10) и филогенетичните мрежи на хаплогрупи I-M423, E-V13, R-M458, R-L23* и J-M241 (Фигура 4) в българската и други евразийски популации може да се използват за проследяване на разпространението на тези хаплогрупи във времето и пространството.

Хаплогрупа I-M423

Мрежата на хаплогрупа I-M423 има звездоподобна форма. В центъра й се намира най-честият хаплотип, присъстващ във всички балкански популации. Топологията на мрежата и оценките на възрастта й са в съответствие с предишни проучвания [Battaglia et al, 2009]. Според авторите, Хг I-M423 е генетичен запис на балканските ловци-събирачи от мезолита и на тяхната експанзия след въвеждането на земеделието. Трябва да се отбележи, че българските проби са пръснати из цялата мрежа, като принадлежат към разпространени, редки и единични хаплотипове, което свидетелства за древността на Хг I-M423 в България.

Хаплогрупа E-V13

Филогенетичната мрежа, която свързва хаплотиповете в Хг E-V13 е звездоподобна, излъчваща се от централен хаплотип застъпен предимно в балканските популации. Това устройство на филогенетичната мрежа и оценките на възрастта показват, че Хг E-V13 е претърпяла скорошна и бърза експанзия на Балканите. Най-древната възраст на Хг E-V13 на Балканите датира от мезолита и е установена в Западна България (9.3 ± 3 kya). От резултатите сме изключили босненските хървати и северните гърци, поради големите стандартни грешки. Възрастта на Хг E-V13 за Западна България е подобна на тази за Турция (10.6 ± 3 kya), което показва, че хаплогрупата е присъствала (ако не и произлязла) през мезолита в Западна България, откъдето се разпространила с преминаването към земеделието.

Хаплогрупа R-M458

Мрежата на Y-STR хаплотиповете асоциирани с Хг R-M458 се характеризира със звездоподобен център (отразяващ експанзия) и няколко сложни преплитания (които се дължат на ниската филогенетична резолюция на хаплогрупата). Болшинството от M458 хаплотиповете заемат места в звездоподобната част на мрежата, което е в съгласие с твърдението на [Underhill et al., 2010], че Хг R-M458 е произлязла в Северна-Централна Европа, откъдето впоследствие се е разпространила. Тези демографски събития са отразени и в оценките на възрастта въпреки, че не може да се изключи влиянието на все още неоткрити под-линии в тази хаплогрупа. Следователно древната възраст на Хг R-M458 в Източна България (12.4 ± 5 kya) и Кавказ (10.1 ± 3 kya) трябва да се интерпретира внимателно, поради това, че оценката на възрастта се повишава вследствие редица демографски епизоди.

*Хаплогрупа R-L23**

Мрежата на Хг R-L23* се характеризира с множество преплитания, които потвърждават, че тази хаплогрупа включва подразклонения, които предстои да бъдат открити. Разпределението на честотата и дисперсията на Хг R-L23* и разликите във възрастта й, сочат че най-древното присъствие на тази линия е в района на Черно море, където са установени сходни времеви оценки от следледниковия период (16.8 ± 7 kya в Източна България; 14.3 ± 1 kya в Румъния; 14.0 ± 3 kya в Кавказ и 13.6 ± 2 kya в Анатолия). Не трябва да се правят прибързани заключения относно датирането за Източна България, тъй като значителна част от тази стойност се дължи на доста по-различен единичен хаплотип, чието изключване значително намалява възрастовата оценка (до 9.3 ± 4 kya).

Хаплогрупа J-M241

Хаплогрупа J-M241 показва мрежа чийто централен и най-чест хаплотип е широко разпространен в южните Балкани, което вероятно е следствие на бърза експанзия, започнала през неолита от Мала Азия. Периферията на мрежата е основно заета от хаплотипове установени извън този регион - пулийци, индийци и непалци. Възрастовите оценки на тези популации са стари, което не отговаря на тяхната позиция в мрежата на Хг J-M241. Поради това несъответствие (което може да се дължи на повтарящи се миграции от различни райони) пулийците, индийците и непалците не се вземат предвид при проследяването на произхода и разпространението на Хг J-M241.

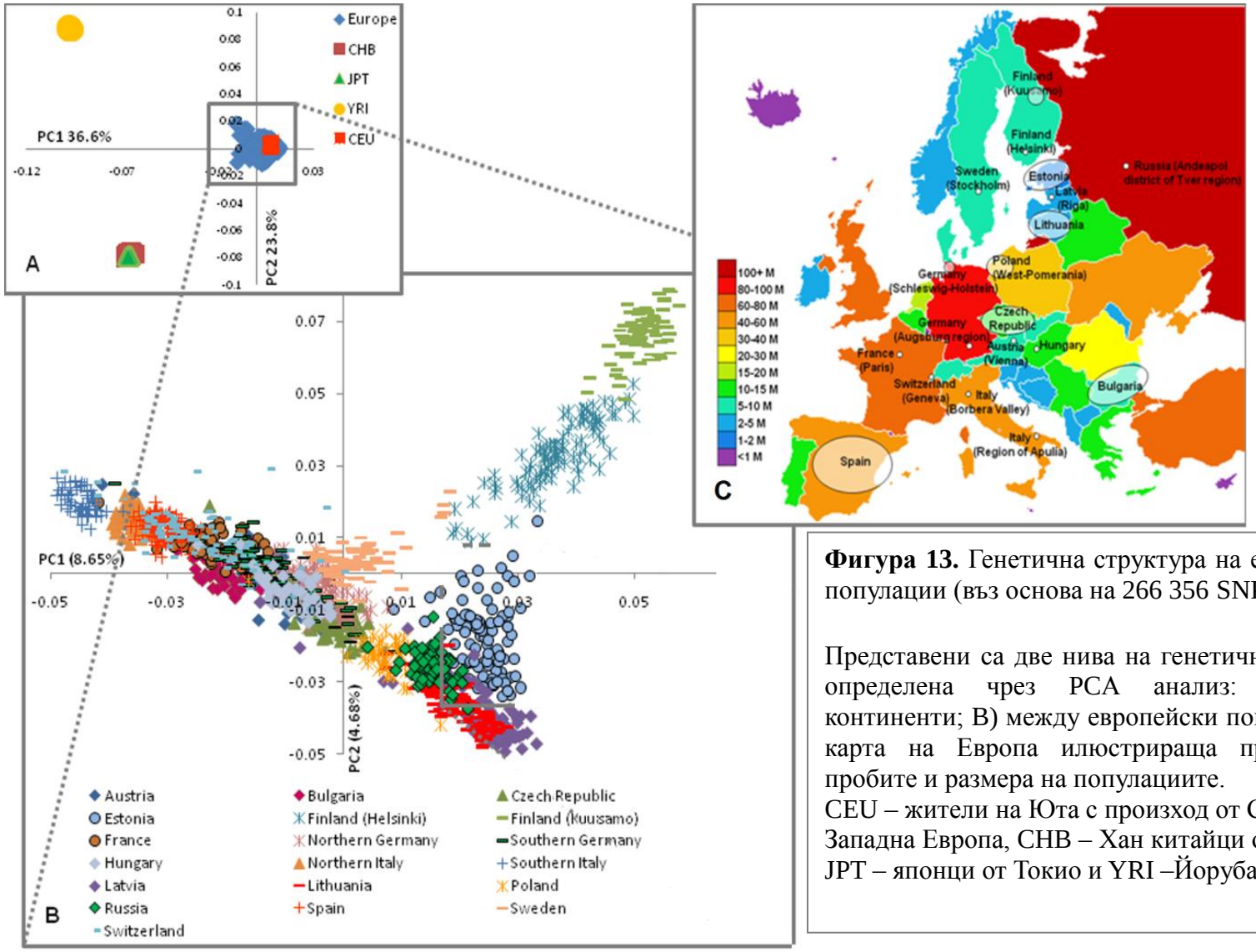
Най-високи стойности за възрастта на хаплогрупата, съвместими с неолитната експанзия, са получени за районите около Черно море, а именно Анатолия (9.1 ± 2 kya) и България - особено за централната й част (7.8 ± 3 kya). Следователно, в тези райони, хаплогрупа J-M241 може да се счита за генетичен сигнал на експанзията на неолитните земеделци към Югоизточна Европа, която вероятно е била усилена чрез скъсването на преградата при Босфора и потопа на Черно море (тогава Понтийско Езеро) с морска вода.

Обсъждане на резултатите от геномния анализ на SNPs

Резултатите от европейското проучване на над 270 000 SNPs са използвани за съставяне на карта на генетичната структура на Европа, която може да улесни междупопулационните генетични изследвания, включително и цялостни-геномни асоциативни проучвания.

При илюстрирането на генетичната структура на европейците е проведен PCA анализ върху 266 356 от изследваните SNPs при европейски и неевропейски популации. В PCA анализа включващ популации от различни континенти (Фигура 13A), първият и вторият главен компонент (PC1 и PC2) обхващат съответно 36.6% и 23.8% от генетичната дисперсия. В PCA анализа на европейските популации (Фигура 13B), първият и вторият главен компонент описват съответно 8.7% и 4.9% от генетичната информация. Те показват, че генетичната структура на европейските популации корелира с географското им положение и че посоката на разпространение на генетичната изменчивост в Европа е с градиент от северозапад към югоизток.

Изследването на цялостната геномна структура в европейски контекст поставя България сред европейските страни, което е в съгласие с междупопулационните сравнения на честотата на мтДНК и Y-хромозомните хаплогрупи.



Фигура 13. Генетична структура на европейските популации (въз основа на 266 356 SNPs).

Представени са две нива на генетична структура, определена чрез PCA анализ: A) между континенти; B) между европейски популации и C) карта на Европа илюстрираща произхода на пробите и размера на популациите.
CEU – жители на Юта с произход от Северна и Западна Европа, CHB – Хан китайци от Пекин, JPT – японци от Токио и YRI –Йоруба от Нигерия.

ЗАКЛЮЧЕНИЕ

Най-мощните средства за проследяване на генетичния произход по майчина и по бащина линия са изследванията на изменчивостта на мтДНК и Y-хромозомата, съответно. До настоящото проучване, научните данни за изменчивостта на еднородителските системи при българите бяха частични, остарели и недостатъчни за характеризиране на генетичния статус и произход на българската популация.

За да запълним липсата на данни за цялостно характеризиране на съвременния български Y-хромозомен и мтДНК генофонд и за да наредим България до европейските страни с подробно описана генетична структура, проведохме всеобхватно и задълбочено проучване на хаплотипове и хаплогрупи в мтДНК и Y-хромозомата и допринесохме за определянето на позицията на българите в генетичната карта на Европа основана на цялостната геномна изменчивост.

Честотата на установените хаплогрупи показва, че както мтДНК, така и Y-хромозомният генофонд на съвременните българи е съставен предимно от евразийски хаплогрупи. Разпределението на тези хаплогрупи на територията на страната е равномерно. Степента на структурираност на Y-хромозомната изменчивост е по-висока между Западна, Централна и Източна България, отколкото между физико-географските райони на страната (Черноморското крайбрежие, югозападния район, Дунавската равнина и Тракия), което показва, че Стара планина най-вероятно не е действала като силна бариера за генен поток.

Анализът на микросателитната изменчивост на Y-хромозомните хаплогрупи разкрива следните интересни резултати: (а) хаплогрупа R-L23* в Източна България датира от следледниковия период; (б) хаплогрупа E-V13 (която вероятно е с южноазиатски произход), присъства в България от мезолита и е започнала да се разпространява след пристигането на неолитното земеделие и (в) хаплогрупа J-M241 най-вероятно отразява неолитната експанзия на земеделците от най-ранните центрове на земеделието по Черноморието.

Установените честоти на хаплогрупите са използвани и за изобразяване на родствените връзки на българите с евразийски и африкански популации. В графиката от анализа на честотата на мтДНК хаплогрупите, българите заемат позиция между източноевропейските и средиземноморските популации. Средиземноморското наследство може да бъде приписано на траките, докато източният принос вероятно се дължи на древните българи и славяните. По отношение на Y-хромозомния хаплогрупен

профил, българите са локализирани между европейците и далече от алтайските популации и говорещите тюркски езици средноазиатски популации. Позицията на българите сред европейските популации се потвърждава и в картата на генетичната структура на Европа според 266 356 SNPs.

Най-важното заключение от настоящото проучване произлиза от генетичната позиция на българите сред другите популации и пренебрежимо малката честота (под 2%) на хаплогрупите типични за алтайските и говорещите тюркски езици средноазиатски популации. Имайки предвид последните исторически изследвания, които показват, че древните българи имат съществен принос при формирането на съвременния български народ, следва че те нямат общ генетичен произход с алтайските популации и говорещите тюркски езици средноазиатски популации.

Генетичният принос на различните популации при формирането на съвременния български генофонд може пряко да се определи чрез филогенетичен анализ на ДНК, възстановена от костни останки открити на територията на страната. Освен това, за целите на фармакогеномиката и проучванията на честите болести би било полезно да се проведе и подобно по обхват и мащаб проучване на генетичната структура на целия геном на българите.

ИЗВОДИ

1. Определени са вариантите и честотите на мтДНК хаплогрупи и хаплотипове в съвременната българска популация.

2. Доказано е, че българският митохондриален генофонд е почти изцяло съставен от хаплогрупи, типични за западноевразийските популации. Общатата честота на източноазиатски (M, C, D и Z) и африкански (L0, L3 и M1) мтДНК хаплогрупи е много ниска (1.3%).

3. Установено е, че българската мтДНК изменчивост е хомогенна за страната.

4. Определена е позицията на българската популация сред западноевразийски и северноафрикански популации въз основа на мтДНК хаплогрупните профили. Българите заемат междинна позиция между източноевропейските и средиземноморските популации. Допуснато е, че средиземноморското наследство може да бъде приписано на траките, докато източният принос вероятно се дължи на древните българи и славяните.

5. Установени са честотите на Y-хромозомни хаплогрупи при съвременните българи.

6. Доказано е, че българският Y-хромозомен генофонд е почти изцяло представен от западноевразийски хаплогрупи. Генетичните приноси от Централна Азия (Хг C-M217), Северна Евразия (Хг N-M231) и Югозападна Азия (производни на Хг Q-M242, Хг L-M61 и Хг R-M124) са незначителни (общо 1.5%).

7.Установено е, че българският Y-хромозомен генофонд е доста еднороден в страната и се предполага, че Стара планина не е действала като силна бариера за генен поток.

8. Доказано е чрез микросателитен анализ на Y-хромозомните хаплогрупи, че Хг R-L23* в Източна България датира от следледниковия период; Хг E-V13 присъства в България от мезолита и започнала да се разпространява след пристигането на неолитното земеделие; а Хг J-M241 най-вероятно отразява неолитната експанзия на земеделците от най-ранните центрове на земеделието по Черноморието.

9. Честотата на Y-хромозомните хаплогрупи, определя позицията на българите в европейския клъстер, далече от алтайските популации и от популациите живеещи на север от региона на Памир.

10. Приведени са доказателства към хипотезата, че древните българи (които според последните исторически изследвания имат съществен принос при формирането на съвременния български народ) нямат общ генетичен произход с алтайските популации и говорещите тюркски езици средноазиатски популации.

НАУЧНИ ПРИНОСИ

1. За първи път всеобхватно е определена генетичната структура на българите по Y-хромозомата и мтДНК и резултатите са анализирани в микро- и макрогеографски мащаб.

2. Определена е давността и начина на разпространение на чести Y-хромозомни хаплогрупи.

3. Създадена е съществена база за сравнение на съвременните българи с други популации и с древните обитатели на българските земи.

4. За първи път България е поставена върху геномната карта на Европа.

5. Доказано е, че древните българи, които според последните исторически изследвания имат съществен принос при формирането на съвременния български народ, нямат общ генетичен произход с алтайските популации и говорещите тюркски езици средноазиатски популации.

ЛИТЕРАТУРНИ ИЗТОЧНИЦИ

1. Йорданов, София, Проф. Марин Дринов, 2006.
2. Achilli et al., 2007 Am J Hum Genet 80(4): 759-768.
3. Al-Zahery et al., 2011 BMC Evol Biol 11: 288.
4. Alonso et al., 2005 Eur J Hum Genet 13(12): 1293-1302.
5. Alvarez-Iglesias et al., 2009 PLoS One 4(4): e5112.
6. Arredi et al., 2004 Am J Hum Genet 75(2): 338-345.
7. Babalini et al., 2005 Eur J Hum Genet 13(8): 902-912.
8. Balanovsky et al., 2008 Am J Hum Genet 82(1): 236-250.
9. Bandelt et al, 1999 Mol Biol Evol 16(1): 37-48.
10. Bandelt et al, 1995 Genetics 141(2): 743-753.
11. Battaglia et al., 2009 Eur J Hum Genet 17(6): 820-830.
12. Brandstatter et al., 2007 Ann Hum Genet 71(Pt 6): 791-803.
13. Brandstatter et al., 2006 J Legal Med 120(5): 310-314.
14. Brandstatter et al., 2008 BMC Evol Biol 8: 191.
15. Cadenas et al., 2008 Eur J Hum Genet 16(3): 374-386.
16. Capelli et al., 2007 Mol Phylogenet Evol 44(1): 228-239.
17. Cinnioglu et al., 2004 Hum Genet 114(2): 127-148.
18. Coudray et al., 2009 Ann Hum Genet 73(2): 196-214.
19. Cruciani et al., 2007 Mol Biol Evol 24(6): 1300-1311.
20. Derenko et al., 2007 Am J Hum Genet 81(5): 1025-1041.
21. Dulik et al., 2011 PLoS One 6(3): e17548.
22. Flores et al., 2005 J Hum Genet 50(9): 435-441.
23. Fornarino et al., 2009 BMC Evol Biol 9: 154.
24. Garcia et al., 2011 Heredity 106(1): 37-45.
25. Gayden et al., 2007 Am J Hum Genet 80(5): 884-894.
26. Gonzalez et al., 2006 BMC Genomics 7: 124.
27. Hassan H et al., 2008 Am J Phys Anthropol 137(3): 316-323.
28. Hedman et al., 2007 Forensic Sci Int 172(2-3): 171-178.
29. Irwin et al., 2008 Int J Legal Med 122(1): 87-89.
30. Irwin et al., 2010 Int J Legal Med 124(3): 195-204.
31. Karachanak et al., 2013 PLoS One 8(3): e56779.
32. Karafet et al., 2001 Am J Hum Genet. 69(3): 615-628.
33. Karlsson et al., 2006 Eur J Hum Genet. 14(8): 963-970.
34. Lacau et al., 2012 Eur J Hum Genet 20(10): 1063-1070.
35. López-Parra et al., 2009 Ann Hum Genet 73(1): 42-53.
36. Luis et al., 2004 Am J Hum Genet 74(3): 532-544.
37. Lutz et al., 1998 Int J Legal Med 111(2): 67-77.
38. Martinez-Cruz et al., 2012 PLoS One 7(7): e41803.
39. Myres et al., 2011 Eur J Hum Genet 19(1): 95-101.
40. Nonaka et al., 2007 Ann Hum Genet 71(Pt 4): 480-495.
41. Palanichamy et al., 2010 BMC Evol Biol 10: 304.
42. Regueiro et al., 2006 Hum Hered 61(3): 132-143.
43. Regueiro et al., 2012 Gene 498(1): 59-67.
44. Richard et al., 2007 Ann Hum Biol 34(1): 68-79.
45. Richards et al., 2000 Am J Hum Genet 67(5): 1251-1276.
46. Richards et al., 2002 Am J Hum Genet 71(5): 1168-1174.
47. Roostalu et al., 2007 Mol Biol Evol 24(2): 436-448.
48. Saunier et al., 2009 Forensic Sci Int Genet 3(3): e97-103.

49. Semino et al., 2000 Science 290(5494): 1155-1159.
50. Semino et al., 2002 Am J Hum Genet 70(1): 265-268.
51. Sengupta et al., 2006 Am J Hum Genet 78(2): 202-221.
52. Soares et al., 2009 Am J Hum Genet 84(6): 740-759.
53. Soares et al., 2010 Curr Biol 20(4): R174-183.
54. Stevanovitch et al., 2004 Ann Hum Genet 68(Pt 1): 23-39.
55. Tetzlaff et al., 2007 Forensic Sci Int 172(2-3): 218-224.
56. Torroni et al., 2001 Am J Hum Genet 69(4): 844-852.
57. Underhill et al., 2010 Eur J Hum Genet 18(4): 479-484.
58. van Oven and Kayser, 2009 Hum Mutat 30(2): E386-394.
59. Wells et al., 2001 Proc Natl Acad Sci U S A. 98(18): 10244-10249.
60. Zhivotovsky et al., 2006 Mol Biol Evol 23(12): 2268-2270.
61. Zhong et al., 2011 Mol Biol Evol 28(1): 717-727.

НАУЧНИ ПУБЛИКАЦИИ И СЪОБЩЕНИЯ ПО ТЕМАТА

Научни публикации във връзка с дисертационния труд:

1. **Karachanak S**, Grugni V, Fornarino S, Nesheva D, Al-Zahery N, Battaglia V, Carossa V, Yordanov Y, Torroni A, Galabov AS, Toncheva D, Semino O. Y-Chromosome Diversity in Modern Bulgarians: New Clues about Their Ancestry. *PLoS ONE* 2013; 8(3):e56779.

2. **Karachanak S**, Carossa V, Nesheva D, Olivieri A, Pala M, Kashani BH, Grugni V, Battaglia V, Achilli A, Yordanov Y, Galabov AS, Semino O, Toncheva D, Torroni A. Bulgarians *vs* the other European populations: a mitochondrial DNA perspective. *Int J Legal Med* 2012; 126(4):497-503.

3. **Карачанак С**, Нешева Д, Тончева Д. Популационна генетика. Изменчивост и еволюция на генофонда. В: Тончева Д, Медицинска генетика в постгеномната ера. София: Симелпрес; 2010, с.160-173.

4. Nelis M, Esko T, Mägi R, Zimprich F, Zimprich A, Toncheva D, **Karachanak S**, Piskácková T, Balascák I, Peltonen L, Jakkula E, Rehnström K, Lathrop M, Heath S, Galan P, Schreiber S, Meitinger T, Pfeufer A, Wichmann HE, Melegh B, Polgár N, Toniolo D, Gasparini P, D'Adamo P, Klovins J, Nikitina-Zake L, Kucinskas V, Kasnauskiene J, Lubinski J, Debniak T, Limborska S, Khrunin A, Estivill X, Rabionet R, Marsal S, Julià A, Antonarakis SE, Deutsch S, Borel C, Attar H, Gagnebin M, Macek M, Krawczak M, Remm M, Metspalu A. Genetic structure of Europeans: a view from the North-East. *PLoS One* 2009; 4(5).

ОБЩ ИМПАКТ ФАКТОР: 10.771

Научни съобщения на конгреси във връзка с дисертационния труд:

1. **Karachanak S**, Fornarino S, Grugni V, Semino O, Toncheva D, Galabov AS, Atanasov B. Poster presentation: Y-Chromosome haplogroups in Bulgarians. Bulgarian-Japanese Symposium. Genomics and Proteomics in Personalized Medicine 19 - 20 March 2009, Sofia, Bulgaria, p. 37, P1.14.

2. **Karachanak S**, Fornarino S, Grugni V, Semino O, Toncheva D, Galabov AS, Atanasov B. Poster presentation: Major mtDNA Haplogroups in Bulgarians. Bulgarian-Japanese Symposium. Genomics and Proteomics in Personalized Medicine 19 - 20 March 2009; Sofia, Bulgaria, p. 38, P1.15.

3. **Karachanak SK**, Carossa V, Nesheva D, Achilli A, Olivieri A, Yordanov Y, Galabov A, Semino O, Torroni A, Toncheva D. Poster presentation: Genetic variation in Bulgarians: a mitochondrial DNA perspective. European Human Genetics Conference 12 – 15 June 2010, Gothenburg, Sweden, p. 260: P10.39.

4. **Karachanak S**, Grugni V, Nesheva D, Battaglia V, Fornarino S, Al-Zahery N, Carossa V, Yordanov Y, Galabov A, Atanasov B, Torroni A, Toncheva D, Semino O. Poster presentation: Y-Chromosome genetic variation of modern Bulgarians. European Human Genetics Conference 28 - 31 May 2011, Amsterdam, The Netherlands, p. 346: P10.80.

5. **Karachanak S**, Grugni V, Nesheva D, Al-Zahery N, Battaglia V, Nici C, Carossa V, Yordanov Y, A. Torroni A, Galabov AS, Semino O, Toncheva D. Poster presentation: A Y -chromosome portrait of modern Bulgarians as viewed from different spatiotemporal aspects. European Human Genetics Conference 23 - 26 June 2012, Nürnberg, Germany, p 260: P10.58.

6. **Karachanak S**, Galabov A, Nesheva D, Jordanov J, Semino O, Torroni A, Toncheva D. Oral presentation: Genetic structure of modern Bulgarians. Bulgarian-Japanese Symposium. Genomics and Proteomics in Personalized Medicine 08 December 2012, Sofia, Bulgaria.

БЛАГОДАРНОСТИ

Изказвам най-сърдечни благодарности на:

-научния си ръководител проф. д-р Драга Тончева, за гласуваното ми доверие, предоставената ми възможност за обучение и цялостната й подкрепа и помощ при изработването, написването и оформянето на настоящата дисертация;

-на акад. проф. Ангел Гълъбов за безценните съвети и препоръки при обсъждането на резултатите при написването на публикациите и настоящата дисертация;

-на проф. Орнелла Семино и проф. Антонио Торрони от Катедрата по биология и биотехнологии "L. Spallanzani" в Университета в Павия, Италия за съветите и насоките при генотипирането и статистическия анализ на експерименталните резултати;

-на проф. генерал Стоян Тонев – началник на Военно Медицинска Академия (ВМА), на ръководителя на Центъра по трансфузионна хематология д-р Румен Попов и на екипа на Центъра за събирането на участници за проучването;

-на проф. Петър Добрев за помощта при тълкуването на резултатите;

-на чл.-кор. Йордан Йорданов за помощта при организирането на проучването;

-на д-р Мая Кичева – управител на Проген ООД, София за съдействието при анализа на микросателити на Y-хромозомата;

-на колегите от Катедрата по Медицинска генетика за помощта при изработването и оформянето на настоящата дисертация;

-на колегите от Катедрата по биология и биотехнологии "L. Spallanzani" към Университета в Павия, Италия за помощта при генотипирането и статистическата обработка на резултатите;

-на всички донори на кръвни проби за доверието;

-на моите родители, сестра ми, съпруга ми, дъщеря ми и на д-р Ягода Василева за подкрепата.

SUMMARY

The most powerful tools for tracing matrilineal and patrilineal population ancestry are the studies of mitochondrial DNA (mtDNA) and Y-chromosome diversity, respectively. The previous mtDNA and Y-chromosome studies of Bulgarians were partial, outdated and insufficient to characterize their genetic composition and origin. In order to fill this gap in the European population genetics, we have performed a large and comprehensive study of the genetic structure of the Bulgarian population from mtDNA and Y-chromosome perspective and we have also contributed to the determination of the position of Bulgarians in the genetic map of Europe based on whole genome diversity.

The analysis of the Bulgarian mtDNA pool was performed on a sample comprising 855 males, whose parents are born in different provinces of the country. The Y-chromosome diversity was evaluated on 808 individuals, which are part of the particicpants in the mtDNA analysis. The whole genome analysis of more than 270 000 autosomal single nucleotide polymorphisms (SNPs) was performed on a different sample comprising 47 Bulgarian males and females.

The mtDNA genotyping was performed by sequencing of ~750 bp of the control region, followed by restriction fragment analysis of markers in the coding region to confirm mtDNA haplogroup (Hg) or subhaplogroup affiliation. In the Y-chromosome survey, we have analyzed 75 biallelic markers in hierarchical order (1 marker was analyzed by amplicon size detection, 4 by sequencing, 38 by restriction fragment analysis and 32 markers by denaturing high performance liquid chromatography). Part of the samples belonging to informative haplogroups were analyzed for 17 Y-STR markers.

The obtained results show that the Bulgarian mitochondrial pool is composed almost completely by haplogroups typical of West Eurasian populations. The presence of East Asian (M, C, D and Z) and African (L0, L3 and M1) mtDNA haplogroups is negligible (1.3%).

The Bulgarian mitochondrial diversity is homogenuous across the country. In the Principal Component Analysis (PCA) Bulgarians locate in an intermediate position between Eastern European and Mediterranean populations. The Mediterranean legacy could be attributed to the Thracians, whereas the Eastern contribution in likely due to the proto-Bulgarians and the Slavs.

The Bulgarian Y-chromosome composition is primarily represented by Western Eurasian haplogroups. Genetic contributions from Central Asia (Hg C-M217), Northern Eurasia (Hg N-M231) and South-West Asia (Hg Q-M242 derivatives, L-M61 and R-M124) were detected at almost negligible frequencies (1.5%).

The Bulgarian Y-chromosome composition is rather homogenous across the country and it shows that the Balkan Mountains probably did not act as a strong barrier to gene flow.

The analysis of Y-STR variation within haplogroups provided the following interesting results: Hg R-L23* in Eastern Bulgaria dates from the post-last glacial period; Hg E-V13 is present in Bulgaria since the Mesolithic, from where it expanded after the arrival of farming; and Hg J-M241 probably reflects the Neolithic westward expansion of farmers from the earliest sites along the Black Sea.

In the PCA analysis of Y-chromosome haplogroup frequencies Bulgarians group with European populations, apart form Altaic populations and the populations residing in the north of the Pamir region. The position of Bulgarians among other European populations is further confirmed in the PCA analysis based on whole genome SNP data.

The haplogroup profile of Bulgarians and their position among other populations in the PCA analyses suggest that a common ancestry between the proto-Bulgarians (which according to the most recent historical studies made a substabtial contribution to the modern Bulgarian people) and the Altaic and Central Asian Turkic-speaking populations either did not exist or was negligible.

In the present study the genetic structure of Bulgarians from mtDNA and Y-chromosome perspective is comprehensively described for the first time. The obtained results, which represent a substantial basis for comparison with other populations, are analyzed on a macro- and micro-geographic levels. Taking into considerations the most recent historical studies, which indicate a substantial proto-Bulgarian input to the contemporary Bulgarian people, the observed patterns reinforce the hypothesis that a common genetic ancestry between the proto-Bulgarians and the Altaic and Central Asian Turkic-speaking populations either did not exist or was negligible.